

Quantitative Security of Block Ciphers: Designs and Cryptanalysis Tools

THÈSE N° 4208 (2008)

PRÉSENTÉE LE 14 NOVEMBRE 2008

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE SÉCURITÉ ET DE CRYPTOGRAPHIE
PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Thomas BAINÈRES

ingénieur en systèmes de communication EPF
de nationalités française et suisse et originaire de Vira (Gambarogno) (TI)

acceptée sur proposition du jury:

Prof. A. Lenstra, président du jury
Prof. S. Vaudenay, directeur de thèse
Dr H. Gilbert, rapporteur
Prof. S. Morgenthaler, rapporteur
Prof. J. Stern, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2008

Contents

I	An Introduction to Modern Cryptology and an Approach to the Design and Cryptanalysis of Block Ciphers	1
1	Shannon’s Theory of Secrecy	3
1.1	The Encryption Model: Preserving Confidentiality	3
1.2	Perfect Secrecy and the Vernam Cipher	5
1.3	Going Beyond Perfect Secrecy	6
1.4	Thesis Outline	6
2	Computationally Bounded Adversaries	9
2.1	Black Box Attacks: Determining the Secret Key Length	9
2.2	New Directions in Cryptography: reducing Confidentiality to Authenticity	11
3	Block Ciphers Design: a Top-Down Approach	13
3.1	Iterated Block Ciphers and Key Schedules	13
3.2	Round Functions Based on Feistel Schemes	14
3.3	Round Functions Based on Lai-Massey Schemes	15
3.4	Round Functions Based on Substitution-Permutation Networks	15
3.5	Providing Diffusion: on the Need for Multipermutations	16
3.6	Providing Confusion: Mixing key bits	17
3.7	The Advanced Encryption Standard	17
4	The Luby-Rackoff Model: Statistical Attacks against Block Ciphers	19
4.1	The Perfect Cipher and Security Models	19
4.2	From Distinguishing to Key Recovery	20
4.3	Linear Cryptanalysis	22
5	Notations and Elementary Results	25
5.1	Random Variables, Probabilities, Strings, etc.	25
5.2	Vector Norms and Fundamental Inequalities	26
5.3	Asymptotic Notations	27

II	On the (In)Security of Block Ciphers: Tools for Security Analysis	29
6	Distinguishers Between Two Sources	31
6.1	A Typical Introduction to Simple Hypothesis Testing	31
6.2	An Alternate View through the Method of Types	33
6.3	The Best Distinguisher: an Optimal Solution	36
6.4	The Best Distinguisher: Data Complexity Analysis	39
6.5	The Best Distinguisher: Examples and Pathological Distributions	48
6.6	The Best Distinguisher: Case where the Distributions are Close to Each Other	50
6.7	The Best Distinguisher: Case where one of the Distributions is Uniform	53
6.8	The Best Distinguisher: Case where one Hypothesis is Composite	54
6.9	A General Heuristic Method to Compute the Advantage of an Arbitrary Distinguisher	56
6.10	Case where One of the Distributions is Unknown: the <i>Squared</i> Distinguisher Family	58
7	Projection-Based Distinguishers Between two Sources	67
7.1	On the Need for New Distinguishers	67
7.2	Best Distinguisher made Practical Using Compression	68
7.3	Linear Distinguishers for Binary Sources	70
7.4	Links between Best, Projection-Based, and Linear Distinguishers for Bi- nary Sources	72
7.5	Extending the Notion of Linear Probability to Arbitrary Sets	81
7.6	Linear Distinguishers for Sources over Arbitrary Sets	84
7.7	A Fundamental Link Between Projection-Based and Linear Distinguishers	87
7.8	Links with Differential Cryptanalysis	92
8	Projection-Based Distinguishers Between two Oracles	95
8.1	From Random Sources to Random Oracles	95
8.2	Cryptanalysis Complexity by means of Transition and Bias Matrices	97
8.3	Piling-up Transition Matrices	101
8.4	Generalized Linear Cryptanalysis of Block Ciphers	105
8.5	The Block Cipher DEAN: a Toy Example for our Generalization of Linear Cryptanalysis	113
8.6	A \mathbf{Z}_{100}^{16} Generalized Linear Cryptanalysis of TOY100	115
9	A Generalized Linear Cryptanalysis of SAFER K/SK	119
9.1	The SAFER Family	120
9.2	Linear Cryptanalysis of SAFER: from \mathbf{Z}_2^8 to \mathbf{Z}_{2^8}	123
9.3	Attacks on Reduced-Round Versions of SAFER	128
9.4	Implementation of the Attack on 2 Rounds	132
9.5	Conclusion	135

III	Block Cipher Designs and Security Proofs	137
10	Provable Security and the Decorrelation Theory	139
10.1	The Luby-Rackoff Model	141
10.2	Computing the Advantage by means of Distribution Matrices	142
10.3	From Linear Cryptanalysis and Differential Cryptanalysis to other Iterated Attacks	147
10.4	Decorrelation of Feistel Ciphers	150
10.5	Decorrelation Modules: Avoiding Algebraic Constructions	152
11	Dial C for Cipher	157
11.1	A Description of the Block Cipher C	157
11.2	Exact Security against 2 -limited Adversaries	161
11.3	Consequences for Iterated Attacks of Order 1, Linear and Differential Cryptanalysis	168
11.4	Exact Security against Linear and Differential Cryptanalysis	169
11.5	Towards the Perfect Cipher	174
11.6	Provable Security against Impossible Differentials	175
11.7	Taking the Key-Schedule into Account	177
11.8	Unproved Security against other Attacks	180
11.9	A Fast Variant of C without Security Compromise	181
11.10	Implementation and Performances	182
11.11	Summary	184
12	KFC: the Krazy Feistel Cipher	185
12.1	From the SPN of C to the Feistel Network of KFC	186
12.2	A Good Round Function for the Feistel Scheme	186
12.3	Exact Security of F_{KFC} against 2 -limited Adversaries	189
12.4	Bounding the Security of F_{KFC} against Adversaries of Higher Order	194
12.5	KFC in Practice	201
12.6	Further Improvements	202
13	Conclusion and Future Work	203
IV	Appendixes	205
A	A Proof of Sanov's Theorem	207
B	Proof of Lemma 6.6	211
C	Proofs of the Lemmas Used in Example 7.3	215
D	The Substitution Box of DEAN27.	217

E	Complementary Informations on SAFER	219
5.1	List of Some of the Possible Successions of Patterns on the Linear Layer	219
5.2	Sequences of Three Weights	224
5.3	Complexities of the Attacks against 3, 4, and 5 Rounds	224

Abstract

Block ciphers probably figure in the list of the most important cryptographic primitives. Although they are used for many different purposes, their essential goal is to ensure confidentiality. This thesis is concerned by their *quantitative security*, that is, by *measurable attributes* that reflect their ability to guarantee this confidentiality.

The first part of this thesis deals with well know results. Starting with Shannon's Theory of Secrecy, we move to practical implications for block ciphers, recall the main schemes on which nowadays block ciphers are based, and introduce the Luby-Rackoff security model. We describe distinguishing attacks and key-recovery attacks against block ciphers and show how to turn the firsts into the seconds. As an illustration, we recall linear cryptanalysis which is a classical example of statistical cryptanalysis.

In the second part, we consider the (in)security of block ciphers against statistical cryptanalytic attacks and develop some tools to perform optimal attacks and quantify their efficiency. We start with a simple setting in which the adversary has to distinguish between two sources of randomness and show how an optimal strategy can be derived in certain cases. We proceed with the practical situation where the cardinality of the sample space is too large for the optimal strategy to be implemented and show how this naturally leads to the concept of *projection-based distinguishers*, which reduce the sample space by compressing the samples. Within this setting, we re-consider the particular case of linear distinguishers and generalize them to sets of arbitrary cardinality. We show how these distinguishers between random sources can be turned into distinguishers between random oracles (or block ciphers) and how, in this setting, one can generalize linear cryptanalysis to Abelian groups. As a proof of concept, we show how to break the block cipher TOY100, introduce the block cipher DEAN which encrypts blocks of decimal digits, and apply the theory to the SAFER block cipher family.

In the last part of this thesis, we introduce two new constructions. We start by recalling some essential notions about provable security for block ciphers and about Serge Vaudenay's Decorrelation Theory, and introduce new simple modules for which we prove essential properties that we will later use in our designs. We then present the block cipher C and prove that it is immune against a wide range of cryptanalytic attacks. In particular, we compute the *exact* advantage of the best distinguisher limited to two plaintext/ciphertext samples between C and the perfect cipher and use it to compute the exact value of the maximum expected linear probability (resp. differential probability) of C which is known to be inversely proportional to the number of samples

required by the best possible linear (resp. differential) attack. We then introduce KFC a block cipher which builds upon the same foundations as C but for which we can prove results for higher order adversaries. We conclude both discussions about C and KFC by implementation considerations.

Keywords: Cryptography, block cipher, statistical cryptanalysis, linear cryptanalysis, hypothesis testing, SAFER, Decorrelation Theory

Résumé

Les algorithmes de chiffrement à clef secrète font très certainement partie des primitives cryptographiques les plus importantes. Bien qu'ils soient utilisés à des fins très diverses, leur principale fonction est d'assurer la confidentialité des données. Cette thèse s'intéresse à leur *sécurité quantitative*, c'est-à-dire aux *attributs mesurables* qui reflètent leur habilité à garantir cette confidentialité.

La première partie de cette thèse traite d'un certain nombre de résultats bien connus. En partant de la théorie du secret de Shannon, nous considérons les implications pratiques pour les algorithmes de chiffrement à clef secrète, nous rappelons les schémas élémentaires sur lesquels ces derniers sont conçus, et introduisons le modèle de Luby et Rackoff. Nous décrivons les attaques visant à distinguer une permutation aléatoire d'une autre puis les attaques dont l'objectif est de retrouver la clef secrète pour enfin montrer comment les premières peuvent entraîner les deuxièmes. En guise d'exemple, nous rappelons les concepts de la cryptanalyse linéaire qui est un exemple classique de cryptanalyse statistique.

Dans la deuxième partie, nous considérons l'(in)sécurité des algorithmes de chiffrement à clef secrète face aux attaques cryptanalytiques statistiques et développons quelques outils pour exécuter certaines attaques et quantifier leur efficacité. Nous considérons un cadre initial très simple dans lequel un adversaire doit distinguer une source aléatoire d'une autre et montrons que, dans certains cas, une stratégie optimale peut être trouvée. Nous traitons ensuite le cas pratique dans lequel la cardinalité de l'espace échantillon est trop grande pour que la stratégie optimale puisse être utilisée telle quelle, ce qui entraîne naturellement la définition de *distingueurs basés sur des projections* qui réduisent l'espace en compressant chaque échantillon. Dans cette optique, nous reconsidérons le cas des distingueurs linéaires et les généralisons aux ensembles de cardinalité arbitraire. Nous montrons comment ces distingueurs entre des sources aléatoires peuvent être transformés en distingueurs entre des oracles aléatoires et comment, de cette façon, il est possible de généraliser la cryptanalyse linéaire aux groupes Abéliens. En guise de preuve de concept, nous montrons comment casser l'algorithme de chiffrement TOY100, introduisons l'algorithme DEAN qui permet de chiffrer des blocs de chiffres décimaux, et appliquons la théorie à la famille d'algorithmes SAFER.

Dans la dernière partie de cette thèse, nous proposons deux nouvelles constructions. Nous commençons par rappeler quelques notions essentielles concernant la sécurité prouvée des algorithmes de chiffrement à clef secrète et la Théorie de la Décorrélation développée par Serge Vaudenay. Nous introduisons de nouveaux modules

pour lesquels un certain nombre de résultats de sécurité peuvent être prouvés et qui seront au coeur des deux constructions à suivre. Nous présentons ensuite l'algorithme de chiffrement C et prouvons sa sécurité contre un certain nombre d'attaques. En particulier, nous calculons l'avantage *exact* du meilleur distingueur limité à deux paires de textes clairs/chiffrés entre C et l'algorithme de chiffrement parfait et utilisons ce résultat pour calculer la valeur exacte de la valeur moyenne maximum de la probabilité linéaire (ainsi que celle de la valeur moyenne de la probabilité différentielle) de C que l'on sait être inversement proportionnelle au nombre d'échantillons nécessaires pour mener une attaque concluante. Nous introduisons ensuite KFC , un algorithme qui repose sur les mêmes bases que C mais pour lequel nous arrivons à prouver des résultats concernant des adversaires d'ordres plus élevés. Dans les deux cas, nous concluons la discussion par des considérations expérimentales.

Mots-clefs: Cryptographie, algorithme de chiffrement à clef secrète, cryptanalyse statistique, cryptanalyse linéaire, test d'hypothèse, SAFER, Théorie de la Décorrélation