"Provable" Security Against Differential and Linear Cryptanalysis

Kaisa Nyberg

Aalto University School of Science and Nokia, Finland
kaisa.nyberg@aalto.fi

Abstract. In this invited talk, a brief survey on the developments of countermeasures against differential and linear cryptanalysis methods is presented.

1 Nonlinearity of S-boxes

Throughout the eighties the unpublished design criteria of the DES had inspired various authors to invent formal nonlinearity criteria for S-boxes such as the *strict avalanche criterion* [30] and the *propagation criterion* [27]. At the same time, correlation attacks on combination generators inspired definitions of *correlation immunity* [29] and *perfect nonlinearity* [21] of Boolean functions. W. Meier and O. Staffelbach realized that perfect nonlinear Boolean functions had been invented before under the name *bent functions* [28,12]. Then the discovery of the differential cryptanalysis method [4] lead to the notion of *perfect nonlinear S-boxes* [22], with the property that for any non-zero input difference the output differences are uniformly distributed. In particular, the output difference zero would occur with the same probability as the non-zero output differences and would significantly improve the probability of the two-round iterative characteristic for a Feistel cipher as pointed out to the author by E. Biham at Eurocrypt 1991. It also means that perfect nonlinear S-boxes cannot be bijective, even worse, the number of input bits must be at least twice the number of output bits [22].

It was clear that the requirement of perfect nonlinearity must be relaxed. But it was not sufficient to take care that the output bits were highly nonlinear Boolean functions as in [26], but also all non-zero linear combinations of the output bits should be highly nonlinear as noted in [23], where the definition of nonlinearity of a vector Boolean function was formulated. The importance of nonlinearity as a cryptographic criterion was highlighted even more as the linear cryptanalysis method was presented by M. Matsui in 1993 [20]. The relationship between nonlinearity (resistance against linear cryptanalysis) and differential uniformity (resistance against differential cryptanalysis) was established in [8]. Since then H. Dobbertin and C. Carlet followed by many other authors have contributed with combinatorial designs and constructions that are almost perfect nonlinear (APN) or satisfy other nonlinearity criteria of S-boxes.

2 CRADIC

We observed that if the differential probabilities of a round function of a Feistel cipher are bounded from above, then also the differential probabilities over four rounds of the cipher are bounded by a significantly smaller bound. There is a penalty of allowing zero output difference as noted by E. Biham, but it takes only one more round to achieve the same security level. In [25] we formulated and proved the following theorem.

Theorem 1. (KN Theorem) *It is assumed that in a DES-like cipher with* $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$ *the round keys are independent and uniformly random. Then the probability of an* s*-round differential,* $s \geq 4$*, is less than or equal to* $2p_{max}^2$*.*

Here

$$\begin{aligned} p_{\max} &= \max_{\beta} \max_{\alpha_R \neq 0} \Pr[\alpha_L + f(E(X + \alpha_R)) + K) + f(E(X) + K) = \beta_R] \\ &\leq p_f = \max_{b} \max_{a \neq 0} \Pr[f(Y + a) + f(Y) = b] \end{aligned}$$

If f bijective, then the claim of the KN Theorem holds for $s \geq 3$, in which case the multiplier 2 can be removed [1].

The high nonlinearity of the Cube function $f(x) = x^3$ in $\mathbb{F}_{2^n}$ had been observed already in [26]. It is bijective for odd n only, so we made one-bit adjustments to it, so that it was possible to fit it into a balanced $2(n-1)$-bit Feistel cipher as a round function. We called this cipher CRADIC, as Cipher Resistant Against DIfferential Cryptanalysis, but in public it became known as KN Cipher. The cipher was later broken using algebraic cryptanalysis making use of the low degree of the Cube monomial.

Since then, designers of block ciphers continue using small nonlinear S-boxes in the spirit of C. Shannon. Would it be possible to use a monolithic algebraic construction? Recently, the Discrete Logarithm function was proved to achieve optimum algebraic immunity [7]. Let α generator of the multiplicative group $\mathbb{F}_{2^n}^*$ and set

$$f(x) = \begin{cases} \log_\alpha(x), & \text{for } x \neq 0 \\ (1, 1, \ldots, 1,) & \text{for } x = 0. \end{cases}$$

Then f gives an n-bit S-box. Previously, it is known that any single output bit of f exhibits asymptotically low correlation with linear functions [6]. The correlations are bounded from above by

$$\mathcal{O}(n\, 2^{-n/2}).$$

But no useful general upper bound is known to the linearity of combinations of output bits. The known bounds increase exponentially as the length of the linear mask grows [7,14]. Later we managed to establish a smaller bound where the increase is exponential with respect to the number of output bits involved, that is, the Hamming weight of the mask. In experiments, however, it seems that the linearity does not grow exponentially but essentially slower. It remains an open question, whether CRADIC would be secure if the Cube function were replaced by the Discrete Logarithm function.

3 Linear Hulls

The essential notion in the KN Theorem is *differential* first introduced in [18]. The approach taken in this work was to model an iterated block cipher as a stochastic process and assume that the rounds are independent. This can be achieved for a key-alternating cipher by selecting the round keys to be statistically independent and then taking the average over all keys. Under the *hypothesis of stochastic equivalence* it is then possible to draw conclusions about the behaviour of the cipher for a fixed unknown key. We adopted the same stochastic model and introduced in [24] the concept of linear hull and proved the following result for the expected squared correlation.

Theorem 2. (Linear Hull Theorem) *Let X, K and Y be random variables in $\mathbb{F}_2^m$, $\mathbb{F}_2^\ell$, and $\mathbb{F}_2^n$, resp. where $Y = F(X, K)$ and X and K are independent. If K is uniformly distributed, then for all $a \in \mathbb{F}_2^m$ and $b \in \mathbb{F}_2^n$,*

$$\mathrm{Exp}_K \mathrm{corr}(a \cdot X + b \cdot Y)^2 = \sum_{c \in \mathbb{F}_2^\ell} \mathrm{corr}(a \cdot X + b \cdot Y + c \cdot K)^2.$$

Here, for random variable Z in $\mathcal{Z}$ (binary strings) we defined

$$\mathrm{corr}(u \cdot Z) = \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} \Pr[z](-1)^{u \cdot z}. \tag{1}$$

Then the linear hull (originally called as approximate linear hull) was defined as the set of all linear approximations

$$ALH(a, b) = \{a \cdot X + b \cdot Y + c \cdot K \,|\, c \in \mathbb{F}_2^\ell\}$$

of plaintext, ciphertext and key, with fixed input and output masks a and b, but letting the key mask vary. Thus taking squares of the correlations and summing over c gives the average correlation over the cipher with plaintext mask a and ciphertext mask b.

J. Daemen abandoned the Markov cipher model and took the fixed key approach [11]. He investigated correlations of linear approximations over a key alternating block cipher E, with round functions $x \mapsto f_i(x + K_i)$, and fixed set of round keys $K_0, \ldots, K_r$. Given vector Boolean function: $f : \mathbb{F}_2^n \mapsto \mathbb{F}_2^m$ with $f = (f_1, \ldots, f_m)$,, where $b \cdot f$ are Boolean functions, for all $b \in \mathbb{F}_2^m$, the correlation between $b \cdot f(x)$ and $a \cdot x$ is defined by

$$c_f(a, b) = \frac{1}{2^n} \left(\#\{x \in \mathbb{F}_2^n \,|\, b \cdot f(x) = a \cdot x\} - \#\{x \in \mathbb{F}_2^n \,|\, b \cdot f(x) \neq a \cdot x\}\right)$$

Then the correlation of a composed function computed as the matrix product

$$c_{f \circ g}(a, b) = \sum_u c_g(a, u) c_f(u, b),$$

from where we obtain

$$c_E(u_0, u_r) = \sum_{u_1,\ldots,u_{r-1}} (-1)^{u_0 \cdot K_0 + \ldots + u_r \cdot K_r} \prod_{i=1}^{r} c_{f_i}(u_{i-1}, u_i),$$

where u_0 and u_r are the linear masks of data after 0 and r rounds of encryption, respectively. This result holds for all fixed keys. By taking the squares and averaging over uniformly distributed and independent keys we get as a corollary

$$\text{Average}_{K_0,\ldots,K_r} c_E(u_0, u_r)^2 = \sum_{u_1,\ldots,u_{r-1}} \prod_{i=1}^{r} c_{f_i}(u_{i-1}, u_i)^2.$$

This result was given in [24] for the special case of DES. Related to this, let us also observe that the correlation of a single trail of a linear hull, taken over plaintext, ciphertext and key, gives another presentation of the piling-up lemma

$$\text{corr}(a \cdot X + b \cdot Y + c \cdot K) = \prod_{i=1}^{r} c_{f_i}(u_{i-1}, u_i),$$

where $a = u_0$, $b = u_r$, and c is in unique correspondence with the trail masks $u_1, \ldots, u_{r-1}$.

Finally let us make an observation of the effect of key scheduling, which should be designed in such a way that the magnitudes of the correlations

$$c_E(u_0, u_r) = \sum_{u_1,\ldots,u_{r-1}} (-1)^{u_0 \cdot K_0 + \ldots + u_r \cdot K_r} \prod_{i=1}^{r} c_{f_i}(u_{i-1}, u_i)$$

do not vary too much with the key. This can be achieved if all dominating trail correlations are of about equal magnitude and the map:

$$(u_1, \ldots, u_{r-1}) \mapsto sign\left(\prod_{i=1}^{r} c_{f_i}(u_{i-1}, u_i)\right)$$

is highly nonlinear. Then the correlations $|c_E(u_0, u_r)|$ are bounded by the small linearity bound. Known examples of mappings with highly nonlinear correlation sign functions are bent functions and the Cube function. For bent functions the sign function is also bent. For the Cube function, correlations are zero in a half space while restricted in the other half space the sign function is bent.

4 Provable Security in Practice

It would be easier to achieve security guarantees against differential and linear attacks for round functions composed of a highly nonlinear monolithic design. In case of substitution permutation networks and similar designs such as AES, cryptographers must work harder. The basic approach is to design the diffusion layer in such a way that the minimum number of active S-boxes involved in the attack is large enough to make the linear trail correlations and differential

characteristic probabilities sufficiently small. To achive this goal, the designers of the AES used MDS matrices for creating larger S-boxes and the Wide-Trail Strategy for ensuring diffusion of trails over the entire width of the cipher [10]. Then obtaining any useful upper bounds to linear correlations and differential probabilities becomes hard. The best known upperbounds for 4 and more rounds are due to L. Keliher [16].

The block cipher PRESENT makes use of bit permutations between rounds for optimal diffusion [5]. Its hardware optimized S-box exhibits, however, strong linear correlations for single-bit masks. Consequently, fairly accurate estimates of correlations can be obtained using single-bit linear approximation trails. As demonstrated in [9], linear hull effect is significant and therefore linear attacks are more powerful than initially estimated by the designers. The other side of the coin is that now we have better estimates of resistance of PRESENT against linear attacks. Can the linear hull effect for PRESENT be computed with sufficient accuracy using single-bit trails only is an interesting question.

5 Linear Approximations and Distributions

The correspondence between correlations of linear projections and probability distributions has been well-known for cryptographers since at least [2] but not exploited in cryptanalysis until in the multidimensional linear cryptanalysis [15]. It allows to use a number of linear approximations simultaneously. More generally, let Z be a vector of (binary) random variables over domain $\mathcal{Z}$. By applying the inverse Walsh-Hadamard transform to (1) we get

$$p_z = \Pr[z] = \sum_{u \in \mathcal{Z}} \mathrm{corr}(u \cdot Z)(-1)^{u \cdot z}.$$

In cryptanalysis, Z is a random variable, which can be sampled from cipher data, such as multidimensional linear approximation, difference, or ciphertext from chosen biased plaintext, anything expected to have non-random behaviour. In this sense, linear approximations, that is, linear projection $z \mapsto u \cdot z$ gives a universal tool for analyzing probability distributions. For example, G. Leander used it to prove that the statistical saturation attack averaged over the fixations and the multidimensional linear cryptanalysis attack are essentially the same [19].

This approach is not restricted to binary variables but can be extended to any finite group. For example, projections $x \mapsto ux \bmod p$, for p prime, have been used in cryptanalysis of block ciphers with non-binary diffusion layer [3]. This leads to the following generalized notion of correlation

$$c_f(u, w) = \frac{1}{q} \sum_{x \in \mathbb{Z}_q} e^{\frac{2\pi i}{p} w f(x)} e^{-\frac{2\pi i}{q} ux}$$

for a function $f : \mathbb{Z}_q \to \mathbb{Z}_p$ and positive integers p and q. The generalized bent functions achieve the smallest linearity with respect to generalized correlation [17]. The Discrete Logarithm function for integers is another example

with known asymptotic upper bound of linearity [13]. This upperbound is of the same magnitude than the bound conjectured to the binary Discrete Logarithm function.

Given such Z related to a cipher, how many samples of Z is needed to distinguish it from a true random variable? If the distribution of Z is close to uniform, then the answer can be given in terms of the capacity of the distribution Z defined as follows:

$$C(Z) = M \sum_{z \in \mathcal{Z}} (p_z - \frac{1}{M})^2,$$

where $M = |\mathcal{Z}|$. Using the relationship between the distribution and correlations we obtain

$$C(Z) = \sum_{u \neq 0} |\mathrm{corr}(u \cdot Z)|^2.$$

Let us summarize the known upper bounds of data complexities for two commonly used distinguishers.

The strongest distinguisher based on the log-likelihod ratio (LLR) requires good knowlege of the probability distribution of Z. If it is available, then the data requirement of the LLR distinguisher can be given as:

$$N_{\mathrm{LLR}} = \frac{\lambda}{C(Z)},$$

where the constant λ depends only on the success probability.

In cryptanalysis, the variable Z and its probability distribution typically depend on the unknown key. While the χ^2 distinguisher is less optimal than the LLR, it can be used also in this case, as it does not require knowledge of the distribution of Z. Its data requirement is

$$N_{\chi^2} = \frac{\lambda' \sqrt{M}}{C(Z)}, \quad \text{where}$$

$$\lambda' \approx (\sqrt{2} + 2) \Phi^{-1}(P_S) \approx \lambda.$$

Cryptanalysts aim at minimizing the data complexity. To be able to use the LLR bound, they must make convincing arguments that LLR works. Else they are left with the higher value given by the χ^2 complexity bound. Cryptographers want to work in the opposite direction and claim as high values as possible for the data complexity. In general, provable security may be difficult to achieve given only such upper bounds of average data complexities. It takes practical experiments and other evidence to see what the actual distinguishing data complexities are and how much they vary with the keys.

Acknowledgement

Thanks to Céline and Risto for their help in the final editing of the paper.

References

1. Aoki, K.: On maximum non-averaged differential probability. In: Tavares, S.E., Meijer, H. (eds.) Selected Areas in Cryptography '98, SAC'98, Proceedings. LNCS, vol. 1556, pp. 118–130. Springer (1999)
2. Baignères, T., Junod, P., Vaudenay, S.: How Far Can We Go Beyond Linear Cryptanalysis? In: Lee, P.J. (ed.) Advances in Cryptology – ASIACRYPT '04, Proceedings. LNCS, vol. 3329, pp. 432–450. Springer (2004)
3. Baignères, T., Stern, J., Vaudenay, S.: Linear cryptanalysis of non binary ciphers. In: Adams, C.M., Miri, A., Wiener, M.J. (eds.) Selected Areas in Cryptography, 14th International Workshop, SAC 2007, Revised Selected Papers. LNCS, vol. 4876, pp. 184–211. Springer (2007)
4. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. In: Menezes, A., Vanstone, S.A. (eds.) Advances in Cryptology - CRYPTO '90, Proceedings. LNCS, vol. 537, pp. 2–21. Springer (1991)
5. Bogdanov, A., Knudsen, L., Leander, G., Paar, C., Poschmann, A., Robshaw, M., Seurin, Y., Vikkelsoe, C.: PRESENT: An Ultra-Lightweight Block Cipher. In: Paillier, P., Verbauwhede, I. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2007, Proceedings. LNCS, vol. 4727, pp. 450–466. Springer (2007)
6. Brandstätter, N., Lange, T., Winterhof, A.: On the non-linearity and sparsity of Boolean functions related to the discrete logarithm in finite fields of characteristic two. In: WCC 2005 Proceedings. LNCS, vol. 3969, p. 135143. Springer (2006)
7. Carlet, C., Feng, K.: An infinite class of balanced vectorial boolean functions with optimum algebraic immunity and good nonlinearity. In: IWCC 2009 Proceedings. LNCS, vol. 5557, pp. 1–11. Springer (2009)
8. Chabaud, F., Vaudenay, S.: Links between differential and linear cryptoanalysis. In: Santis, A.D. (ed.) Advances in Cryptology - EUROCRYPT'94, Proceedings. LNCS, vol. 950, pp. 356–365. Springer (1995)
9. Cho, J.Y.: Linear cryptanalysis of reduced-round PRESENT. In: Pieprzyk, J. (ed.) Topics in Cryptology - CT-RSA 2010, The Cryptographers' Track at the RSA Conference 2010. LNCS, vol. 5985, pp. 302–317. Springer (2010)
10. Daemen, J., Rijmen, V.: The Design of Rijndael – AES, the Advanced Encryption Standard. Springer (2002)
11. Daemen, J., Govaerts, R., Vandewalle, J.: Correlation matrices. In: Preneel, B. (ed.) Fast Software Encryption 1994. LNCS, vol. 1008, pp. 275–285. Springer, Heidelberg (1995)
12. Dillon, J.F.: Elementary Hadamard difference sets. In: Proceedings of the Sixth Southeastern Conference on Cornbinatorics, Graph Theory and Computing, Boca Raton, Florida. Congressus Numerantium, vol. XIV, pp. 237–249. Utilitas Math., Winnipeg, Manitoba (1975)
13. Hakala, R.M.: An upper bound for the linearity of Exponential Welch Costas functions. Finite Fields and Their Applications (to appear), `http://dx.doi.org/10.1016/j.ffa.2012.05.001`
14. Hakala, R.M., Nyberg, K.: On the nonlinearity of discrete logarithm in $\mathbb{F}_{2^n}$. In: Carlet, C., Pott, A. (eds.) Sequences and Their Applications – SETA 2010. LNCS, vol. 6338, pp. 333–345. Springer (2010)
15. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In: Mu, Y., Susilo, W., Seberry, J. (eds.) Information Security and Privacy, 13th Australasian Conference, ACISP 2008, Proceedings. LNCS, vol. 5107, pp. 203–215. Springer (2008)

16. Keliher, L.: Refined analysis of bounds related to linear and differential cryptanalysis for the AES. In: Fourth Conference on the Advanced Encryption Standard (AES4). LNCS, vol. 3373, pp. 42–57. Springer, Heidelberg (2005)
17. Kumar, P.V., Scholtz, R.A., Welch, L.R.: Generalized bent functions and their properties. J. Combin. Theory Ser. A 40(1), 90–107 (1985)
18. Lai, X., Massey, J., Murphy, S.: Markov ciphers and differential cryptanalysis. In: Davies, D.W. (ed.) Advances in Cryptology – EUROCRYPT 1991, Proceedings. LNCS, vol. 547, pp. 141–152. Springer (1991)
19. Leander, G.: On linear hulls, statistical saturation attacks, present and a cryptanalysis of puffin. In: Paterson, K.G. (ed.) Advances in Cryptology - EUROCRYPT 2011, Proceedings. LNCS, vol. 6632, pp. 303–322. Springer (2011)
20. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In: Helleseth, T. (ed.) Advances in Cryptology – EUROCRYPT '93, Proceedings. LNCS, vol. 765, pp. 386–397. Springer (1994)
21. Meier, W., Staffelbach, O.: Nonlinearity criteria for cryptographic functions. In: Advances in Cryptology - EUROCRYPT'89, Proceedings. LNCS, vol. 434, pp. 549–562 (1990)
22. Nyberg, K.: Perfect nonlinear S-boxes. In: Advances in Cryptology - EUROCRYPT'91. LNCS, vol. 547, pp. 378–386. Springer-Verlag (1991)
23. Nyberg, K.: On the construction of highly non-linear permutations. In: Advances in Cryptology - EUROCRYPT'92, Proceedings. LNCS, vol. 658, pp. 92–98. Springer-Verlag (1993)
24. Nyberg, K.: Linear approximation of block ciphers. In: Advances in Cryptology - EUROCRYPT'94, Proceedings. LNCS, vol. 950, pp. 439–444. Springer-Verlag (1995)
25. Nyberg, K., Knudsen, L.R.: Provable security against a differential attack. Journal of Cryptology 8(1), 27–37 (1995)
26. Pieprzyk, J.: On bent permutations. Tech. rep., The University of South Wales, Department of Computer Science (1991), Presented at the International Conference on Finite Fields, Coding Theory and Advances in Communications and Computing, Las Vegas, 1991
27. Preneel, B., Leekwijck, W.V., Linden, L.V., Govaerts, R., Vandewalle, J.: Propagation characteristics of Boolean functions. In: Advances in Cryptology - EUROCRYPT '90, Proceedings. LNCS, vol. 473, pp. 161–173. Springer (1991)
28. Rothaus, O.S.: On "bent" functions. J. Combinatorial Theory Ser. A(20), 300–305 (1976)
29. Siegenthaler, T.: Correlation-immunity of nonlinear combining functions for cryptographic applications. IEEE Transactions on Information Theory 30(5), 776–780 (1984)
30. Webster, A.F., Tavares, S.E.: On the design of s-boxes. In: Advances in Cryptology - CRYPTO'85. LNCS, vol. 219, pp. 523–534. Springer-Verlag (1985)