Linear Cryptanalysis of Non Binary Ciphers
with an Application to SAFER

Thomas Baignères*[1], Jacques Stern[2], and Serge Vaudenay[1]

[1] EPFL
CH-1015 Lausanne – Switzerland
thomas.baigneres@epfl.ch, serge.vaudenay@epfl.ch
[2] École normale supérieure
Département d'Informatique 45, rue d'Ulm
75230 Paris Cedex 05, France
jacques.stern@ens.fr

Abstract. In this paper we re-visit distinguishing attacks. We show how to generalize the notion of linear distinguisher to arbitrary sets. Our thesis is that our generalization is the most natural one. We compare it with the one by Granboulan et al. from FSE'06 by showing that we can get sharp estimates of the data complexity and cumulate characteristics in linear hulls. As a proof of concept, we propose a better attack on their toy cipher TOY100 than the one that was originally suggested and we propose the best known plaintext attack on SAFER K/SK so far. This provides new directions to block cipher cryptanalysis even in the binary case. On the constructive side, we introduce DEAN18, a toy cipher which encrypts blocks of 18 decimal digits and we study its security.

1 Introduction and Mathematical Background

In the digital age, information is always seen as a sequence of bits and, naturally, most practical block ciphers and cryptanalytic tools assume that the text space is made of binary strings. In the literature, a block cipher over a finite set M is commonly defined as a set of permutations $\mathsf{C}_k : \mathsf{M} \to \mathsf{M}$ indexed by a key $k \in \mathcal{K}$, with $\mathsf{M} = \{0,1\}^\ell$ [36]. This restriction is quite questionable though, as it is easy to think of specific settings in which it could be desirable to adapt the block size to the data being encrypted. For example, when considering credit card numbers, social security numbers, payment orders, schedules, telegrams, calendars, string of alphabetical characters,... it seems that there is no reason what so ever to restrict to binary strings. Whereas an apparently straightforward solution would be to encode the data prior encryption, the loss in terms of simplicity (inevitably affecting the security analysis) and of efficiency would be unfortunate.

Although most modern block ciphers (e.g., $[1, 2, 9, 21, 48]$) are defined on a binary set, practical and efficient examples of block ciphers defined on a set of arbitrary size exist (see for example Schroeppel's "omnicipher" Hasty Pudding [45]). Some others, although still defined on binary sets, use a mixture of

* Supported by the Swiss National Science Foundation, 200021-107982/1

group laws over the same set. For example, IDEA [30] combines three group structures: exclusive bit or, addition modulo 2^{16} and a tweaked multiplication modulo $2^{16}+1$. Designing a block cipher with an arbitrary block space can be particularly challenging since the state of the art concerning alternate group structures is very limited. Although differential cryptanalysis [5] (through the theory of Markov ciphers [29]) can be specified over an arbitrary group, linear cryptanalysis [34] is based on a metric (the linear probability) that sticks to bit strings. Applying it to a non-binary block cipher would at least require to generalize this notion. Although several generalizations of linear cryptanalysis exist [15–20,23,24,28,37,42,46,47,50], to the best of our knowledge, none easily applies to, say, modulo 10-based block ciphers. So far, only Granboulan et al. [13] provide a sound treatment on non-binary cipher but mostly address differential cryptanalysis. We show that, for linear cryptanalysis, their data complexity cannot be precisely estimated. Furthermore, no cumulating effect of "linear hull" seems possible. We propose another notion of nonlinearity which fixes all those drawbacks and makes us believe that it is the most natural one.

Outline. In the three first sections of this paper, we re-visit distinguishing attacks on random sources (like stream ciphers or pseudo-random generators) and on random permutations (like block ciphers), in the spirit of Baignères et al. [3], but without assuming that domains are vector spaces. Consequently, the only structure we can consider on these sets is that of finite Abelian groups. In particular, we reconsider linear, optimal, and statistical distinguishers against random sources and linear distinguishers against block ciphers.

The following sections apply this theory to TOY100 and SAFER K/SK (on which we devise the best known plaintext attack so far, showing that our generalization can be useful even in the binary case). On the constructive side, we introduce DEAN18, a toy cipher which encrypts blocks of 18 decimal digits.

Notations. Throughout this paper, random variables $X, Y, \ldots$ are denoted by capital letters, whilst their realizations $x \in \mathcal{X}, y \in \mathcal{Y}, \ldots$ are denoted by small letters. The cardinal of a set $\mathcal{X}$ is denoted $|\mathcal{X}|$. The probability function of a random variable X following a distribution D is denoted $\Pr_{X \in_{\mathsf{D}} \mathcal{X}}[x]$, $P_{\mathsf{D}}(x)$, or abusively $\Pr_X[x]$, when the distribution is clear from the context. A sequence $X_1, X_2, \ldots, X_n$ of n random variables is denoted $\boldsymbol{X}^n$. Similarly, a sequence $x_1, x_2, \ldots, x_n$ of realizations is denoted $\boldsymbol{x}^n$. We call *support* of a distribution D the set of all $x \in \mathcal{X}$ such that $P_{\mathsf{D}}(x) \neq 0$. As usual, "iid" means "independent and identically distributed". $\mathbf{1}_{\mathrm{A}}$ is 1 if the predicate A is true, 0 otherwise. The distribution function of the standard normal distribution is denoted

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2} du \ .$$

Mathematical Background. Let G be a finite group of order n. We let $L^2(G)$ denote the n-dimensional vector space of complex-valued functions f on G. The conjugate $\overline{f}$ of f is defined by $\overline{f}(a) = \overline{f(a)}$ for all $a \in \mathsf{G}$. We define an *inner* *product* on $L^2(G)$ by $(f_1, f_2) = \sum_{a\in\mathsf{G}} f_1(a)\overline{f_2}(a)$. The Euclidean norm of $f \in L^2(G)$ is simply $\|f\|_2 = (f, f)^{1/2} = (\sum_a |f(a)|^2)^{1/2}$. Consequently, $L^2(G)$ is a Hilbert Space. A *character* of G is a homomorphism $\chi : \mathsf{G} \to \mathbf{C}^\times$, where $\mathbf{C}^\times$ is the multiplicative group of nonzero complex numbers. Then $\chi(1) = 1$ and $\chi(a_1a_2) = \chi(a_1)\chi(a_2)$ for all $a_1, a_2 \in \mathsf{G}$. Clearly, $\chi(a)$ is a nth root of unity, hence $\overline{\chi}(a) = \chi(a)^{-1}$. The *product* of two characters χ_1 and χ_2 is defined as $\chi_1\chi_2(a) = \chi_1(a)\chi_2(a)$ for all $a \in \mathsf{G}$. The character $\mathbf{1}$ defined by $\mathbf{1}(a) = 1$ for all $a \in \mathsf{G}$ is the neutral element for this operation. Clearly, $\chi^{-1} = \overline{\chi}$. The set $\widehat{\mathsf{G}}$ of all characters of G is the *dual group* of G and is isomorphic to G.

Lemma 1 (Theorems 4.6 and 4.7 in [40]). *Let G be a finite Abelian group of order n, and let $\widehat{\mathsf{G}}$ be its dual group. If $\chi \in \widehat{\mathsf{G}}$ (resp. $a \in \mathsf{G}$) then*

$$\sum_{a\in\mathsf{G}} \chi(a) = \begin{cases} n & \text{if } \chi = \mathbf{1}, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{resp.} \qquad \sum_{\chi\in\widehat{\mathsf{G}}} \chi(a) = \begin{cases} n & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases}$$

If $\chi_1, \chi_2 \in \widehat{\mathsf{G}}$ (resp. $a, b \in \mathsf{G}$) then

$$\sum_{a\in\mathsf{G}} \chi_1(a)\overline{\chi_2}(a) = \begin{cases} n & \text{if } \chi_1 = \chi_2, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{resp.} \qquad \sum_{\chi\in\widehat{\mathsf{G}}} \chi(a)\overline{\chi}(b) = \begin{cases} n & \text{if } a = b, \\ 0 & \text{otherwise.} \end{cases}$$

If $\chi_1, \chi_2 \in \widehat{\mathsf{G}}$, we deduce $(\chi_1, \chi_2) = n$ if $\chi_1 = \chi_2$ and 0 otherwise. Therefore, the n characters of $\widehat{\mathsf{G}}$ is an orthogonal basis of the vector space $L^2(G)$.

Definition 2 (Fourier transform). *The Fourier transform of $f \in L^2(\mathsf{G})$ is the function $\widehat{f} \in L^2(\widehat{\mathsf{G}})$ such that $\widehat{f}(\chi) = (f, \chi)$ for all $\chi \in \widehat{\mathsf{G}}$.*

If $\widehat{f} \in L^2(\widehat{\mathsf{G}})$ is the Fourier transform of $f \in L^2(G)$, the Fourier inversion is

$$f = \frac{1}{n}\sum_{\chi\in\widehat{\mathsf{G}}} \widehat{f}(\chi)\chi.$$

Theorem 3 (Plancherel's formula). *If $\widehat{f} \in L^2(\widehat{\mathsf{G}})$ is the Fourier transform of $f \in L^2(G)$, then $\|\widehat{f}\|_2 = \sqrt{n}\|f\|_2$.*

Consider the particular case where $\mathsf{G} = \{0, 1\}^k$, $\chi_u(a) = (-1)^{u\bullet a}$ for all $u, a \in \mathsf{G}$, and where $\bullet$ denotes the inner dot product in G. The mapping $u \mapsto \chi_u$ is an isomorphism between G and $\widehat{\mathsf{G}}$. Consequently, when $\mathsf{G} = \{0, 1\}^k$ any character χ of G can be expressed as $\chi(a) = (-1)^{u\bullet a}$ for some $u \in \mathsf{G}$. In linear cryptanalysis, u is called a mask and there is a one-to-one mapping between masks and characters. So, it seems reasonable to generalize linear cryptanalysis on any finite Abelian group by using characters instead of masks.

2 Distinguishing a Biased Source of Finite Support

We consider a source generating a sequence of d iid random variables $\boldsymbol{Z}^d$ following a distribution $\mathsf{D_s}$ of finite support $\mathcal{Z}$. We wonder whether $\mathsf{D_s} = \mathsf{U}$ or

$\mathsf{D_s} = \mathsf{D}$ (where U is the uniform distribution over $\mathcal{Z}$) knowing that these two events are equiprobable and that one of them is eventually true. An algorithm which takes a sequence of d realizations $\boldsymbol{z}^d$ as input and outputs either 0 or 1 is a *distinguisher* $\mathcal{D}$ limited to d samples. The ability to distinguish a distribution from another is the *advantage* of the distinguisher and is defined by

$$\mathrm{Adv}^d_{\mathcal{D}} = \Big| \Pr_{\mathsf{U}^d}[\mathcal{D} \text{ outputs } 1] - \Pr_{\mathsf{D}^d}[\mathcal{D} \text{ outputs } 1] \Big|, \tag{1}$$

which is a quantity an adversary would like to maximize. If the set $\mathcal{Z}$ has the structure of an Abelian group, we denote it G and denote by n its cardinality.

2.1 Optimal Distinguishers

Due to the Neyman-Pearson lemma, the best distinguisher is based on the maximum likelihood strategy. It consists in comparing $\Pr_{\mathsf{U}^d}[\boldsymbol{z}^d]$ and $\Pr_{\mathsf{D}^d}[\boldsymbol{z}^d]$.

Definition 4 (Baignères et al. [3]). *The* Squared Euclidean Imbalance *(*SEI*) of a distribution* D *of finite support* $\mathcal{Z}$ *is defined by*

$$\Delta(\mathsf{D}) = |\mathcal{Z}| \sum_{z \in \mathcal{Z}} \left(P_{\mathsf{D}}(z) - \frac{1}{|\mathcal{Z}|} \right)^2 = |\mathcal{Z}| \sum_{z \in \mathcal{Z}} P_{\mathsf{D}}(z)^2 - 1 = |\mathcal{Z}|\, 2^{-H_2(\mathsf{D})} - 1$$

where $H_2(\mathsf{D})$ *is the Rényi entropy of order 2.*

It was shown in [3] that when using d samples $Z_1, \ldots, Z_d \in \mathcal{Z}$ the advantage of the best distinguisher $\mathcal{A}$ is such that

$$\mathrm{Adv}^d_{\mathcal{D}} \approx 1 - 2\Phi(-\sqrt{\lambda}/2), \tag{2}$$

where $\lambda = d \cdot \Delta(\mathsf{D})$. When $\lambda = 1$ we obtain $\mathrm{Adv}^d_{\mathcal{D}} \approx 0.38$. Note also that when $\lambda \ll 1$, the previous equation simplifies to $\mathrm{Adv}^d_{\mathcal{D}} \approx \sqrt{\frac{\lambda}{2\pi}}$, whereas, when $\lambda \gg 1$, it simplifies to $\mathrm{Adv}^d_{\mathcal{D}} \approx 1 - \frac{4e^{-\lambda/8}}{\sqrt{2\pi\lambda}}$. This motivates the rule of thumb that the data complexity for the best distinguisher should be $d \approx 1/\Delta(\mathsf{D})$.

Using Theorem 3, we obtain the following expression for the SEI.

Lemma 5. *Given a distribution* D *whose support is a finite Abelian group* G *of order* n*, we have* $\Delta(\mathsf{D}) = n\|P_{\mathsf{D}} - P_{\mathsf{U}}\|_2^2 = \|\widehat{P}_{\mathsf{D}} - \widehat{P}_{\mathsf{U}}\|_2^2 = \sum_{\chi \in \widehat{\mathsf{G}} \setminus \{\mathbf{1}\}} \big|\widehat{P}_{\mathsf{D}}(\chi)\big|^2$.

2.2 Linear Probabilities

Typically, performing a linear cryptanalysis [34] against a source of bit-strings of length ℓ consists in analyzing one bit of information about each sample z_i, by means of a scalar product between a (fixed) *mask* $u \in \{0,1\}^\ell$. By measuring the statistical bias of this bit, it is sometimes possible to infer whether $\mathsf{D_s} = \mathsf{U}$ (in which case, the bias should be close to 0) or $\mathsf{D_s} = \mathsf{D}$ (in which case, the bias may be large). Chabaud and Vaudenay [8] adopted the *linear probability* (LP) [35] defined by $\mathrm{LP}_{\mathsf{D}}(u) = (2\Pr_{X \in_{\mathsf{D}} \{0,1\}^\ell}[u \bullet X = 0] - 1)^2 = (\mathrm{E}_{X \in_{\mathsf{D}} \{0,1\}^\ell}((-1)^{u \bullet X}))^2$ as a fundamental measure for linear cryptanalysis. Given the fact that the source is not necessarily binary, it seems natural to generalize the LP as follows.

Definition 6. *For all group character $\chi : \mathsf{G} \to \mathbf{C}^\times$, the* linear probability *of a distribution* D *over* G *with respect to* χ *is defined by*

$$\mathrm{LP}_{\mathsf{D}}(\chi) = \left|\mathrm{E}_{A \in_{\mathsf{D}} \mathsf{G}}\left(\chi(A)\right)\right|^2 = \left|\sum\nolimits_{a \in \mathsf{G}} \chi(a) P_{\mathsf{D}}(a)\right|^2 = \left|\widehat{P_{\mathsf{D}}}(\chi)\right|^2.$$

The LP of χ is simply the square of *magnitude* of the discrete Fourier transform of the probability distribution. In the particular case where $\mathsf{G} = \{0,1\}^\ell$, we can see that for any u we have $\mathrm{LP}_{\mathsf{D}}(u) = \mathrm{LP}_{\mathsf{D}}(\chi_u)$, so that Definition 6 indeed generalizes the notion of linear probability.

Granboulan et al. [13] adopted a different metric which can be expressed by $\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi) = \max_z \left(\Pr_{A \in_{\mathsf{D}} \mathsf{G}}[\chi(A) = z] - \frac{1}{m}\right)^2$ where m is the order of χ. When $m = 2$, we easily obtain $\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi) = 4 \cdot \mathrm{LP}_{\mathsf{D}}(\chi)$ but when $m > 2$, there is no simple relation. Nevertheless, we have $\mathrm{LP}_{\mathsf{D}}(\chi) \le \frac{m^2}{2} \mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi)$ for $m > 2$. This bound is fairly tight since the following distribution reaches $\mathrm{LP}_{\mathsf{D}}(\chi) = \frac{m^2}{4} \mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi)$: we let $\mathsf{G} = \mathbf{Z}_m$ for $m > 2$, $\chi(x) = e^{\frac{2i\pi}{m}x}$, and $P_{\mathsf{D}}(x) = \frac{1}{m} + \varepsilon \times \cos \frac{2\pi x}{m}$. We have $\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi) = \varepsilon^2$ and we can easily compute $\mathrm{LP}_{\mathsf{D}}(\chi) = \frac{m^2}{4}\varepsilon^2$. This shows that our $\mathrm{LP}_{\mathsf{D}}(\chi)$ maybe quite larger than $\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi)$.

2.3 Linear Distinguisher

We construct a linear distinguisher as follows. Let

$$\mathrm{sa}(\boldsymbol{z}^d; \chi) = \frac{1}{d} \sum_{j=1}^{d} \chi(z_j) \qquad \text{and} \qquad \mathrm{lp}(\boldsymbol{z}^d; \chi) = \left|\mathrm{sa}(\boldsymbol{z}^d; \chi)\right|^2.$$

The statistical average $\mathrm{sa}(\boldsymbol{z}^d; \chi)$ over the sample vector $\boldsymbol{z}^d$ can serve for distinguishing U from D. We define the *order* of the linear distinguisher as the order m of χ in $\widehat{\mathsf{G}}$. For example, linear distinguishers of order 2 correspond to classical linear distinguishers. Note that this order must be reasonable so that the implementations can compute the complex number $\mathrm{sa}(\boldsymbol{z}^d; \chi)$.

The law of large numbers tells us that $\mathrm{lp}(\boldsymbol{z}^d; \chi) \xrightarrow[d \to \infty]{} \left|\mathrm{E}_{Z \in_{\mathsf{D}} \mathsf{G}}(\chi(Z))\right|^2 = \mathrm{LP}_{\mathsf{D}}(\chi)$. Informally, when lp is large, it is likely that $\mathsf{D}_{\mathsf{s}} = \mathsf{D}$, whereas when it is close to 0, it is likely that $\mathsf{D}_{\mathsf{s}} = \mathsf{U}$. Consequently, the advantage of a linear distinguisher $\mathcal{D}$ can be defined by optimizing a decision threshold τ, i.e., we have

$$\mathrm{Adv}^d_{\mathcal{D}}(\chi) = \max_{0<\tau<1} \left|\Pr_{\mathsf{U}^d}[\mathrm{lp}(\boldsymbol{Z}^d; \chi) < \tau] - \Pr_{\mathsf{D}^d}[\mathrm{lp}(\boldsymbol{Z}^d; \chi) < \tau]\right|.$$

When the exact distribution of $\chi(Z)$ (for $Z \in_{\mathsf{D}} \mathsf{G}$) on the unit circle is known, one can build a more powerful distinguisher. However, we will later show that the best improvement factor that is achievable is not particularly large, due to the fact that the order of χ must be small. Besides, one only knows in practice that the distribution of $\chi(Z)$ belongs to a set of m possible distributions. For instance, considering (regular) linear cryptanalysis (that is, using characters of order 2), the expected value of the statistical average is $\pm\epsilon$ and thus lies on circle of radius

ϵ. The sign is unknown as it depends on an unknown key. This generalizes to characters of higher order, for which the *exact* value of the mean of the statistical average might allow to know which of the m possible distributions we are dealing with and thus give more information about the key. As for the distinguishing issue, we rather stick to the simpler statistical test based on $\mathrm{lp}(\boldsymbol{Z}^d;\chi)$ only.

The following theorem allows to lower bound the advantage of a linear distinguisher in terms of the linear probability of the source with respect to D.

Theorem 7. *Let* G *be a finite Abelian group and let* $\chi \in \widehat{\mathsf{G}}$. *Using heuristic approximations, the advantage* $\mathrm{Adv}^d_{\mathcal{D}}$ *of a d-limited linear distinguisher* $\mathcal{D}$ *trying to distinguish the uniform distribution* U *from* D *is such that* $\mathrm{Adv}^d_{\mathcal{D}}(\chi) \succeq 1 - 2\cdot e^{-\frac{d}{4}\mathrm{LP}_{\mathsf{D}}(\chi)}$ *(resp.* $\mathrm{Adv}^d_{\mathcal{D}}(\chi) \succeq 1 - 4\cdot\Phi\big(-\frac{1}{2}\sqrt{d\cdot\mathrm{LP}_{\mathsf{D}}(\chi)}\big)$*) for* χ *of order at least 3 (resp. of order 2), when d is large enough and under the heuristic assumption that the covariance matrix of* $\mathrm{lp}(\boldsymbol{Z}^d;\chi)$ *is the same for both distributions.*[3]

Proof. Let m be the order of χ. We denote $\chi(Z_j) = e^{\frac{2i\pi}{m}\theta_j}$ for all $j = 1,\ldots,d$ and let $X_j = \cos(\frac{2\pi}{m}\theta_j)$ and $Y_j = \sin(\frac{2\pi}{m}\theta_j)$, so that

$$\mathrm{lp}(\boldsymbol{Z}^d;\chi) = \big|\tfrac{1}{d}\textstyle\sum_{j=1}^d X_j + i\cdot\tfrac{1}{d}\sum_{j=1}^d Y_j\big|^2 = \big(\tfrac{1}{d}\sum_{j=1}^d X_j\big)^2 + \big(\tfrac{1}{d}\sum_{j=1}^d Y_j\big)^2.$$

The law of large numbers gives $\frac{1}{d}\sum_{j=1}^d X_j + i\cdot\frac{1}{d}\sum_{j=1}^d Y_j \to \mathrm{E}_{Z\in_{\mathsf{D_s}}\mathsf{G}}(\chi(Z))$ when $d\to\infty$. Considering complex numbers as bidimensional vectors, we obtain from the multivariate central limit theorem [11] that the distribution of $\sqrt{d}(\frac{1}{d}\sum_{j=1}^d(X_j+iY_j) - \mathrm{E}_Z(\chi(Z)))$ tends to the bivariate normal distribution with zero expectation and appropriate covariance matrix Σ. We can show that

$$\Sigma = \begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{pmatrix} \text{ for } m\geq 3 \text{ and } \Sigma = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \text{ for } m = 2.$$

We conclude that, when $\mathsf{D_s} = \mathsf{U}$ and $m \geq 3$, the sums $\frac{1}{\sqrt{d}}\sum X_j$ and $\frac{1}{\sqrt{d}}\sum Y_j$ are asymptotically independent and follow a normal distribution with zero expectation and standard deviation equal to $1/\sqrt{2}$. Consequently, $(\frac{\sqrt{2}}{\sqrt{d}}\sum X_j)^2$ and $(\frac{\sqrt{2}}{\sqrt{d}}\sum Y_j)^2$ both follow a chi-square distribution with 1 degree of freedom and $2\cdot d\cdot\mathrm{lp}(\boldsymbol{Z}^d;\chi) = (\frac{\sqrt{2}}{\sqrt{d}}\sum X_j)^2 + (\frac{\sqrt{2}}{\sqrt{d}}\sum Y_j)^2$ follows a chi-square distribution with 2 degrees of freedom [44]. Hence,

$$\Pr_{\mathsf{U}^d}[2\cdot d\cdot\mathrm{lp}(\boldsymbol{Z}^d;\chi) < \alpha] \xrightarrow[d\to\infty]{} \frac{1}{2}\int_0^\alpha e^{-u/2}du = 1 - e^{-\frac{\alpha}{2}}. \quad (3)$$

On the other hand, by making the heuristic approximation that the covariance matrix is the same in the case where $\mathsf{D_s} = \mathsf{D}$, we similarly obtain that

$$\Pr_{\mathsf{D}^d}\big[2\cdot d\cdot\big|\tfrac{1}{d}\textstyle\sum_{j=1}^d(X_j+iY_j) - \mathrm{E}_Z(\chi(Z))\big|^2 < \beta\big] \approx \displaystyle\frac{1}{2}\int_0^\beta e^{-u/2}du = 1 - e^{-\frac{\beta}{2}}. \quad (4)$$

[3] We use the $\succeq$ symbol instead of $\geq$ to emphasize the heuristic assumptions.

Moreover, assuming that $\tau < \mathrm{LP}_{\mathsf{D}}(\chi)$,

$$\Pr_{\mathsf{D}^d}\left[\mathrm{lp}(\boldsymbol{Z}^d;\chi) < \tau\right] \le \Pr_{\mathsf{D}^d}\left[\left|\tfrac{1}{d}\sum_{j=1}^d (X_j+iY_j) - \mathrm{E}_{\mathsf{D}}(\chi(Z))\right|^2 \ge \left(\sqrt{\mathrm{LP}_{\mathsf{D}}(\chi)}-\sqrt{\tau}\right)^2\right]$$

so that for $\alpha \le 2d\tau$ and $\beta \le 2d(\sqrt{\mathrm{LP}_{\mathsf{D}}(\chi)} - \sqrt{\tau})^2$,

$$\mathrm{Adv}^d_{\mathcal{D}}(\chi) \ge \Pr_{\mathsf{U}^d}\left[2\cdot d\cdot \mathrm{lp}(\boldsymbol{Z}^d;\chi)<\alpha\right] - \Pr_{\mathsf{D}^d}\left[2d\cdot\left|\tfrac{1}{d}\sum(X_j+iY_j) - \mathrm{E}_Z(\chi(Z))\right|^2 \ge \beta\right].$$

Using Approximation (4) with $\tau = \frac{1}{4}\mathrm{LP}_{\mathsf{D}}(\chi)$ we obtain for $\alpha \le \frac{d}{2}\mathrm{LP}_{\mathsf{D}}(\chi)$

$$\mathrm{Adv}^d_{\mathcal{D}}(\chi) \ge 2\cdot \Pr_{\mathsf{U}}\left[2\cdot d\cdot \mathrm{lp}(\boldsymbol{Z}^d;\chi) \le \alpha\right] - 1.$$

Taking (3) as a heuristic approximation with $\alpha = \frac{d}{2}\mathrm{LP}_{\mathsf{D}}(\chi)$ leads to the announced result for $m \ge 3$.

For $m = 2$ and $\mathsf{D_s} = \mathsf{U}$, we have that $\frac{1}{\sqrt{d}}\sum_{j=1}^d X_j$ tends towards a standard normal distribution, so that $\left(\frac{1}{\sqrt{d}}\sum_{j=1}^d X_j\right)^2$ tends towards a chi-square distribution. Consequently,

$$\Pr_{\mathsf{U}^d}\left[d\cdot \mathrm{lp}(\boldsymbol{Z}^d;\chi) < \alpha\right] \xrightarrow[d\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_0^\alpha \frac{e^{-x/2}}{\sqrt{x}}\,dx = 1 - 2\Phi(-\sqrt{\alpha}).$$

Similar techniques than in the $m \ge 3$ case lead to the announced result. $\square$

Note that these lower bounds are only useful (otherwise too low) if the number of samples exceeds $\frac{4\ln 2}{\mathrm{LP}_{\mathsf{D}}(\chi)}$ in the large order case and $\frac{2}{\mathrm{LP}_{\mathsf{D}}(\chi)}$ in the order 2 case. For example, when $d = 4/\mathrm{LP}_{\mathsf{D}}(\chi)$ the advantage is greater than 0.26 in the first case and greater than 0.36 in the second. They validate the rule of thumb that the distinguisher works with data complexity $d \approx 1/\mathrm{LP}_{\mathsf{D}}(\chi)$. In contrast, [13] claims without further justification that $1/\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi)$ samples are sufficient to reach a large advantage. It appears that this approximation overestimates the data complexity, actually equal to $1/\Delta(\chi(Z))$, which lies in between $\frac{1}{m^2}(\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi))^{-1}$ (when all values of $\chi(z)$ are biased, like for the distribution example in Section 2.2 for which we have $\mathrm{LP}_{\mathsf{D}}(\chi) = \frac{m^2}{4}\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi))$ and $\frac{1}{2m}(\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi))^{-1}$ (like when only two output values u_1 and u_2 of χ are biased and the others are uniformly distributed, and for which $\mathrm{LP}_{\mathsf{D}}(\chi) = |u_1 - u_2|^2\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi))$. The correct estimate of the data complexity requires more than just the $\mathrm{LP}^{\mathsf{alt}}_{\mathsf{D}}(\chi)$ quantity.

2.4 Case Study: $\mathbf{Z}^r_m$-based Linear Cryptanalysis.

We illustrate the theory with a concrete example, that is, linear cryptanalysis over the additive group $\mathbf{Z}^r_m$. The m^r characters of this group are called *additive characters modulo* m and are the φ^m_a's for $a = (a_1,\ldots,a_r)$ where $a_\ell \in [0, m-1]$ for $\ell = 1,\ldots,r$ defined by $\varphi^m_a(x) = e^{\frac{2\pi i}{m}\sum_{\ell=1}^r a_\ell x_\ell}$ for $x \in \mathbf{Z}^r_m$ (see [40]).

We revisit an example proposed in [3] where a source generating a random variable $X = (X_1,\ldots,X_{n+1}) \in \mathbf{Z}_4^{n+1}$ is considered (n being any large odd integer). When the source follows distribution U, X is uniformly distributed. When the source follows distribution D, $X_1, \ldots, X_n$ are uniformly distributed mutually independent random variables in $\mathbf{Z}_4$ and $X_{n+1} = B + \sum_{\ell=1}^{n} X_\ell$, where B is either 0 or 1 with equal probability and where the addition is performed modulo 4. Considering X as a bitstring of length $2n+2$, it was shown in [3] that $\max_\alpha \mathrm{LP}_{\mathsf{D}}(\varphi_\alpha^2) = 2^{-(n+1)}$ (the max being taken over classical linear masks), which means that the source cannot be distinguished from a perfectly random one using a classical linear distinguisher. On the other hand, let $a = (-1, \ldots, -1, 1) \in \mathbf{Z}_4^{n+1}$ and consider the character φ_a^4 over $\mathbf{Z}_4^{n+1}$. In this case we have $\mathrm{LP}_{\mathsf{D}}(\varphi_a^4) = \left|\mathrm{E}\left(e^{\frac{\pi i}{2}(X_{n+1} - \sum_{\ell=1}^{n} X_\ell)}\right)\right|^2 = \left|\mathrm{E}\left(e^{\frac{\pi i}{2}B}\right)\right|^2 = \frac{1}{2}$. Note that $\mathrm{LP}_{\mathsf{D}}^{\mathsf{alt}}(\varphi_a^4) = \frac{1}{16}$. Theorem 7 suggests that $d = 8$ would be enough for an advantage greater than 0.26. More specifically, the distinguisher can eventually decide that $\mathsf{D_s} = \mathsf{U}$ as soon as $\varphi_a^4(X) \notin \{1, i\}$ (since 1 and i are the only possible values for $\mathsf{D_s} = \mathsf{D}$) for some sample X and that $\mathsf{D_s} = \mathsf{D}$ if all samples X return $\varphi_a^4(X) \in \{1, i\}$. For this distinguisher, $\mathrm{Adv}_{\mathcal{D}}^d = 1 - \frac{1}{2^d}$, so that $d = 1$ is enough to reach an advantage equal to $\frac{1}{2}$. We notice that there can be a huge gap between linear distinguishers of order 2 and linear distinguishers of order 4.

2.5 A Dash of Differential Cryptanalysis

We can consider a natural (see [13]) generalization of the differential probability (DP) and show the link between the LP and the DP (as in [8]). Let $u \in \mathsf{G}$ be an arbitrary group element. The *differential probability* of distribution D over G is defined by $\mathrm{DP}_{\mathsf{D}}(u) = \Pr[A^{-1} \cdot B = u] = \Pr[A \cdot u = B]$, where A and B are independent random variables following the distribution D. We have $\widehat{\mathrm{DP}}_{\mathsf{D}}(\chi) = \mathrm{LP}_{\mathsf{D}}(\chi)$ for any $\chi \in \widehat{\mathsf{G}}$. Indeed, by definition, $\mathrm{LP}_{\mathsf{D}}(\chi) = \mathrm{E}_{\mathsf{D}}(\chi(A))\mathrm{E}_{\mathsf{D}}(\overline{\chi}(B))$, where A and B are independent random variable following distribution D. Successively using the facts that A and B are independent, that the mean is linear, and that χ is a homomorphism, we have for all $u \in \mathsf{G}$

$$\widehat{\mathrm{LP}}_{\mathsf{D}}(u) = \textstyle\sum_{\chi \in \widehat{\mathsf{G}}} \mathrm{E}_{\mathsf{D}}(\chi(A)\overline{\chi}(B))\chi(u) = \mathrm{E}_{\mathsf{D}}\left(\sum_{\chi \in \widehat{\mathsf{G}}} \chi(A \cdot u)\overline{\chi}(B)\right),$$

which is an expression that can be further simplified using Lemma 1, finally leading to $\widehat{\mathrm{LP}}_{\mathsf{D}}(u) = n\mathrm{E}_{\mathsf{D}}\left(\mathbf{1}_{A \cdot u = B}\right) = n\Pr_{\mathsf{D}}[A \cdot u = B] = n\mathrm{DP}_{\mathsf{D}}(u)$. Generalizing the LP as we do in Definition 6 naturally leads to a real duality between linear and differential cryptanalysis. We note that this is not the case when considering the $\mathrm{LP}_{\mathsf{D}}^{\mathsf{alt}}$ measurement suggested in [13].

2.6 Links between Linear and Optimal Distinguishers

Given Lemma 5 and the definition of LP we obtain the following result.

Theorem 8 (Generalization of Proposition 11 in [3]). *Let* D *be a probability distribution of support* G*. The* SEI *of* D *and the linear probability of* D *are related by*

$$\Delta(\mathsf{D}) = \textstyle\sum_{\chi \in \widehat{\mathsf{G}} \setminus \{\mathbf{1}\}} \mathrm{LP}_{\mathsf{D}}(\chi).$$

This equation is pretty insightful when trying to improve linear distinguishers by using the rule of thumb. If there is a character χ such that $\mathrm{LP}_{\mathsf{D}}(\chi)$ overwhelms all other linear probabilities in the previous equation, a single characteristic χ can be used to approximate the linear hull (that is, the cumulative effect of all the characteristics). In that case, one linear distinguisher becomes nearly optimal in term of required number of samples. As another example we can look at the problem of cumulating linear characteristics. In linear cryptanalysis, if we use k independent characteristics of same bias we can best hope to decrease the data complexity by a factor within the order of magnitude of k. This generalizes results by Kaliski and Robshaw [23] and by Biryukov et al. [6].

We can easily deduce useful results for computing the SEI of combinations of independent sources. Namely, for two independent random variables A and B, $\Delta(A+B) \leq \Delta(A)\Delta(B)$ (Piling-up Lemma) and $\Delta(A||B)+1 \leq (\Delta(A)+1)(\Delta(B)+1)$ so $\Delta(A||B)$ is roughly less than $\Delta(A)+\Delta(B)$ (cumulating effect).

Definition 9. *Let* D *be a probability distribution over a group* G *and let* $\mathrm{LP}_{\mathsf{D}}^{\max}$ *be the maximum value of* $\mathrm{LP}_{\mathsf{D}}(\chi)$ *over* $\chi \in \widehat{\mathsf{G}} \setminus \{\mathbf{1}\}$ *of order dividing* m*, i.e.,*

$$\mathrm{LP}_{\mathsf{D}}^{\max}(m) = \max_{\substack{\chi \in \widehat{\mathsf{G}} \setminus \{\mathbf{1}\} \\ \chi^m = \mathbf{1}}} \mathrm{LP}_{\mathsf{D}}(\chi).$$

We note that $\Delta(\mathsf{D})$ does not depend on the group structure whereas $\mathrm{LP}_{\mathsf{D}}^{\max}$ does. We define a metric $\mathrm{LP}_{\mathsf{D}}^{\mathrm{MAX}}$ which does not.

Definition 10. *Let* D *be a probability distribution of support* G *and* $\Diamond$ *denote an arbitrary group operation on* G*. We define*

$$\mathrm{LP}_{\mathsf{D}}^{\mathrm{MAX}}(m) = \max_{\Diamond} \mathrm{LP}_{\mathsf{D}}^{\max}(m).$$

Corollary 11. *Let* D *be a probability distribution whose support is the finite group* G *of order* n*. For the exponent* m *of* G*, we have*

$$\Delta(\mathsf{D}) \leq (n-1) \cdot \mathrm{LP}_{\mathsf{D}}^{\max}(m) \quad \textit{and} \quad \Delta(\mathsf{D}) \leq (n-1) \cdot \mathrm{LP}_{\mathsf{D}}^{\mathrm{MAX}}(m).$$

This result says that the best distinguisher for D has a data complexity at least $n-1$ times less than the one of the best linear distinguisher.

Going back to the distinguisher based on $\chi(Z)$ for a given χ of order m, we assume that the support of distribution D of $\chi(Z)$ matches the range of χ which is a group G of order $n = m$. Assuming that χ is such that $\mathrm{LP}_{\mathsf{D}}(\chi) = \mathrm{LP}_{\mathsf{D}}^{\max}(m)$ we deduce that the best distinguisher between $\chi(Z)$ and a uniformly distributed random variable on its support needs at most m times less data than the linear distinguisher that we proposed.

2.7 Optimal Distinguisher made Practical using Compression

From a computational point of view, the best distinguisher of Section 2.1 cannot be implemented if the order of the group is too large. We consider this situation by denoting H a finite set of large cardinality N and compress the samples using a *projection*

$$h : \mathsf{H} \longrightarrow \mathsf{G},$$

where G is a set of cardinality $n \ll N$. We assume that h is *balanced*. This implies that $n \mid N$. This projection defines, for a random variable $H \in \mathsf{H}$ of distribution $\widetilde{\mathsf{D}_\mathsf{s}}$ (either equal to the uniform distribution $\widetilde{\mathsf{U}}$ or to $\widetilde{\mathsf{D}}$), a random variable $h(H) = Z \in \mathsf{G}$ of distribution D_s (either equal to U or to D). We can easily prove state the following (intuitive) result by using Cauchy's inequality.

Lemma 12 (Projections reduce the imbalance). *Let* H *and* G *be two finite Abelian groups of order* N *and* n *respectively, such that* $n \mid N$. *Let* $h : \mathsf{H} \to \mathsf{G}$ *be a balanced function. Let* $\widetilde{\mathsf{D}}$ *be a probability distribution of support* H *and let* $H \in \mathsf{H}$ *be a random variable following* $\widetilde{\mathsf{D}}$. *Let* D *denote the distribution of* $h(H) \in \mathsf{G}$. *Then* $\Delta(\mathsf{D}) \leq \Delta(\widetilde{\mathsf{D}})$.

The following theorem shows that, in the particular case where the projection is homomorphic, bounding the linear probability of the source is sufficient to bound the advantage of the best distinguisher on the reduced sample space.

Lemma 13 (Generalization of Theorem 13 in [3]). *Let* H *and* G *be two finite Abelian groups of order* N *and* n *respectively, such that* $n \mid N$. *Let* $h : \mathsf{H} \to \mathsf{G}$ *be a surjective* group homomorphism. *Let* $\widetilde{\mathsf{D}}$ *be a probability distribution of support* H *and let* $H \in \mathsf{H}$ *be a random variable following* $\widetilde{\mathsf{D}}$. *Let* D *denote the distribution of* $h(H) \in \mathsf{G}$. *Then* $\Delta(\mathsf{D}) \leq (n-1)\mathrm{LP}^{\max}_{\widetilde{\mathsf{D}}}(n)$.

Proof. From Theorem 8, we have

$$\Delta(\mathsf{D}) = \sum\nolimits_{\chi\in\widehat{\mathsf{G}}\setminus\{\mathbf{1}\}}\mathrm{LP}_{\mathsf{D}}(\chi) = \sum\nolimits_{\chi\in\widehat{\mathsf{G}}\setminus\{\mathbf{1}\}}\mathrm{LP}_{\widetilde{\mathsf{D}}}(\chi\circ h) \leq (n-1)\max_{\chi\in\widehat{\mathsf{G}}\setminus\{\mathbf{1}\}}\mathrm{LP}_{\widetilde{\mathsf{D}}}(\chi\circ h).$$

We note that $\kappa = \chi\circ h$ is a group character of H such that $\kappa^n = \mathbf{1}$. Consequently, $\max_{\chi\in\widehat{\mathsf{G}}\setminus\{\mathbf{1}\}}\mathrm{LP}_{\widetilde{\mathsf{D}}}(\chi\circ h) \leq \max_{\substack{\kappa\in\widehat{\mathsf{H}}\setminus\{\mathbf{1}\}\\ \kappa^n=\mathbf{1}}}\mathrm{LP}_{\widetilde{\mathsf{D}}}(\kappa)$. □

We stress that the previous theorem only applies when the adversary reduces the text space through a group homomorphism, i.e., in a *linear* way. Indeed, there exists practical examples of random sources with a small $\mathrm{LP}^{\max}_{\widetilde{\mathsf{D}_\mathsf{s}}}$ that are significantly broken when the source space is reduced by a (well chosen) non-homomorphic projection (see the example of Section 2.4 with $h(x) = \mathrm{msb}(\varphi_a^4(x))$ and $\mathsf{G} = \mathbf{Z}_2$). Consequently, the previous result tells us nothing about the advantage of an adversary using an arbitrary projection. In what follows we show a security criterion which is *sufficient* to obtain provable security against *any* distinguisher using a balanced projection.

Theorem 14. *Let* H *and* G *be two finite sets of cardinality* N *and* n *respectively, such that* $n \mid N$. *Let* $h : \mathsf{H} \to \mathsf{G}$ *be a* balanced *projection. Let* $\widetilde{\mathsf{D}}$ *be a probability distribution of support* H *and let* $H \in \mathsf{H}$ *be a random variable following* $\widetilde{\mathsf{D}}$. *Let* D *denote the distribution of* $h(H) \in \mathsf{G}$. *Then*

$$\Delta(\mathsf{D}) \leq (n-1)\mathrm{LP}^{\mathrm{MAX}}_{\widetilde{\mathsf{D}}}(n).$$

Proof. We first define an arbitrary group structure on G. Given h, we can easily construct a group structure on H such that h is a homomorphism. The final result then follows from Lemma 13. □

Consequently, assuming there exists an "efficient" projective distinguisher on $\widetilde{\mathsf{D}}$ using a balanced h on a "small" set G, $\Delta(\mathsf{D})$ must be large and n must be small, therefore, $\mathrm{LP}^{\mathrm{MAX}}_{\widetilde{\mathsf{D}}_s}(n)$ is large. Thus, there exists a group structure on H and a character of small order on this group that define an effective linear cryptanalysis: *if we can efficiently distinguish by compressing the samples, we can also do it linearly.* To the best of our knowledge, all widespread block ciphers provably secure against linear cryptanalysis consider in the proof a *specific* group or field structure on the text space. Usually, the more convenient is the one used to actually define the block cipher. Obviously, a potential adversary is not limited to the description considered by the designers. The previous theorem shows that, provided that a known plaintext attack on the block cipher exists, then some *change* to the group structure of the text space is sufficient to perform a successful linear cryptanalysis of the cipher (note that finding the correct group structure might be a non-trivial task). In other words, although the cipher is stated to be provably secure against linear cryptanalysis, it might not be the case when generalizing linear cryptanalysis to other group structures. This is mainly due to the fact that the SEI does not depend on the group structure given to the text space (only the distance of D from the uniform distribution is relevant) whereas the linear probability is a measure that *depends* on the group structure. Consequently, when proving the resistance against linear cryptanalysis, one should ideally bound the value of $\mathrm{LP}^{\mathrm{MAX}}_{\widetilde{\mathsf{D}}}(m)$ and not of $\mathrm{LP}^{\max}_{\widetilde{\mathsf{D}}}(m)$ (as it is currently the case for most block ciphers).

3 Linear Cryptanalysis of Block Ciphers

The theory developed in the previous section can be applied as-is to study the indistinguishability of a pseudo-random sequence (e.g., the output of a stream cipher) from a perfectly random one. We show in this section how it can be adapted to study the security of block ciphers.

3.1 Generalized Linear Cryptanalysis of Block Ciphers

We consider a block cipher defined on a finite group M and an adversary who is given access to a generator $\mathcal{G}$ generating iid random variables $(M, \mathsf{C}(M)) \in \mathsf{M} \times \mathsf{M}$, where M is a uniformly distributed random variable, and where C is a random permutation of M either equal to C_K (a random instance of a block cipher, the randomness coming from the secret key $K \in \mathsf{K}$) or to C^* (the perfect cipher, that is, a uniformly distributed random permutation defined on M). The objective of the adversary is to guess whether $\mathsf{C} = \mathsf{C}_K$ or $\mathsf{C} = \mathsf{C}^*$ (i.e., if the permutation implemented by $\mathcal{G}$ was drawn uniformly at random among the set of permutations defined by the block cipher or among the entire set of permutations of M) after a limited number d of samples $S_i = (M_i, \mathsf{C}(M_i))$ for $i = 1, \ldots, d$. In a classical linear cryptanalysis (i.e., when $\mathsf{M} = \{0,1\}^n$), the adversary would typically run over all plaintext/ciphertext pairs and add the value of $a \bullet M_i \oplus b \bullet \mathsf{C}(M_i)$ to a counter, where a and b are input/output masks defined on the text space. The adversary eventually guesses whether the generator is implementing an instance of the block cipher or not by measuring the bias of the counter with respect to $d/2$. By choosing the masks with care, the bias may be large when $\mathsf{C} = \mathsf{C}_k$ for some key k. In this situation, the linear probability $\mathrm{LP}^{\mathsf{C}_k}(a, b) = (2 \cdot \Pr_M(a \bullet M \oplus b \bullet \mathsf{C}_k(M) = 0) - 1)^2 = \left|\mathrm{E}((-1)^{a \bullet M \oplus b \bullet \mathsf{C}_k(M)})\right|^2$ estimates the efficiency of the attack against C_k. The following definition extends this notion to non-binary linear cryptanalysis.

Definition 15. *Let $\mathsf{C} : \mathsf{M} \to \mathsf{M}$ be a permutation over a finite set M. Let G_1 and G_2 be two group structures over the same set M. For all group characters $\chi \in \widehat{\mathsf{G}}_1$ and $\rho \in \widehat{\mathsf{G}}_2$ the linear probability of C over M with respect to χ and ρ is defined by*

$$\mathrm{LP}^{\mathsf{C}}(\chi, \rho) = \left|\mathrm{E}_{M \in_{\mathrm{U}} \mathsf{M}}\left(\overline{\chi}(M)\rho(\mathsf{C}(M))\right)\right|^2.$$

If C is a random *permutation, we denote the* expected *linear probability by $\mathrm{ELP}^{\mathsf{C}}(\chi, \rho) = \mathrm{E}_{\mathsf{C}}(\mathrm{LP}^{\mathsf{C}}(\chi, \rho))$.*

As direct computation of the linear probability on a realistic instance of a block cipher is not practical, the cryptanalyst typically follows a bottom-up approach, in which he first computes the linear probability of small building blocks of the cipher and then extends the result to the whole construction. In the following section, we study several typical building blocks on which block ciphers are often based. We illustrate our results on a toy cipher in Appendix 6.

3.2 A Toolbox for Linear Cryptanalysis

We can look at a block cipher as a circuit made of building blocks and in which every edge is attached to a specific group. From this point of view, a linear characteristic is a family mapping every edge to one character of the attached group. The building blocks we consider are represented on Figure 1. If χ_1 and χ_2 are characters on G_1 and G_2 respectively, we denote by $\chi_1 \| \chi_2 : \mathsf{G}_1 \times \mathsf{G}_2 \to \mathbf{C}^\times$ the character mapping $(a, b) \in \mathsf{G}_1 \times \mathsf{G}_2$ on $\chi_1(a)\chi_2(b)$. We assume that the cryptanalyst constructs a linear characteristic in a reversed way [4] (i.e., starting from the end of the block cipher towards the beginning), his objective being to

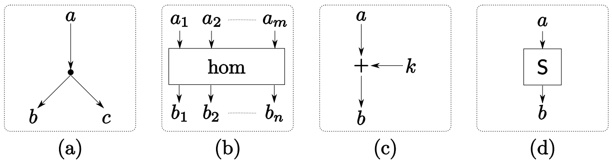

Fig. 1. Typical Building Blocks of Block Ciphers.

carefully choose the characters in order to maximize the linear probability on each individual building block.

Building Block (a): We consider a *duplicate gate* such that $a, b, c \in \mathsf{G}$ and $a = b = c$. Let χ_1, χ_2 be two characters defined over G, we have (by definition) $\chi_1(b)\chi_2(c) = \chi_1(a)\chi_2(a) = \chi_1\chi_2(a)$. Simply denoting (a) the duplicate gate, we have $\mathrm{LP}^{(\mathrm{a})}(\chi_1\chi_2, \chi_1\|\chi_2) = 1$, so that $\chi_1\|\chi_2$ is an appropriate character on the input of the gate.

Building Block (b): We consider a layer that applies a *group homomorphism* from $\mathsf{G} = \mathsf{G}_1 \times \cdots \times \mathsf{G}_m$ to $\mathsf{H} = \mathsf{H}_1 \times \cdots \times \mathsf{H}_n$. We denote the homomorphism by hom, the m inputs as $a_1, a_2, \ldots, a_m$ and the n outputs $b_1, b_2, \ldots, b_n$, so that $\mathsf{hom}(a_1, a_2, \ldots, a_m) = (b_1, b_2, \ldots, b_n)$. Given n characters χ_i on H_i, $i = 1, \ldots, n$, we have $\chi(b_1, \ldots, b_n) = (\chi \circ \mathsf{hom})(a_1, \ldots, a_m)$ for $\chi = \chi_1\| \cdots \|\chi_n$. As $\chi \circ \mathsf{hom}$ is still a homomorphism from G to $\mathbf{C}^\times$ we obtain $\mathrm{LP}^{(\mathrm{b})}(\chi \circ \mathsf{hom}, \chi) = 1$. Note that we do have $\chi \circ \mathsf{hom} = \chi'_1\| \cdots \|\chi'_m$ for some $(\chi'_1, \ldots, \chi'_m) \in \widehat{\mathsf{G}}_1 \times \cdots \times \widehat{\mathsf{G}}_m$, so that χ'_i is an appropriate character for a_i.

Building Block (c): Given $\mathsf{hom}(a) = a+k$ on a given group G (adopting a more traditional additive notation), we have $\chi(b) = \chi(a)\chi(k)$. Since k is constant, $\mathrm{LP}^{(\mathrm{c})}(\chi, \chi) = 1$, so that χ is an appropriate character on the input.

Building Block (d): When considering a (non-homomorphic) permutation S, $\mathrm{LP}^{\mathsf{S}}(\chi, \rho)$ should be computed by considering the substitution table of S.

By piling all relations up on a typical substitution-permutation network C, we obtain a relation of the form $\overline{\chi}(M)\rho(\mathsf{C}(M)) = \left(\prod_i \overline{\chi_i}(X_i)\rho(\mathsf{S}_i(X_i))\right) \times \left(\prod_j \chi_j(k_j)\right)$ where the first product runs over all building blocks of type (d) and the second over building blocks of type (c). Hence, by making the heuristic approximation of independence of all X_i's (which is commonly done in classical linear cryptanalysis), we obtain that

$$\mathrm{LP}^{\mathsf{C}}(\chi, \rho) \approx \textstyle\prod_i \mathrm{LP}^{\mathsf{S}_i}(\chi_i, \rho_i).$$

This is the classical single-path linear characteristic. Provided that we can lower bound (e.g. using branch numbers) the number of active substitution boxes S to b and that we have $\mathrm{LP}^{\mathsf{S}_i}_{\max} \leq \lambda$ for all boxes we obtain that $\mathrm{LP}^{\mathsf{C}}_{\max}$ is heuristically bounded by λ^b for single-path characteristics.

For multipath characteristics, we easily obtain the linear hull effect [41].

Theorem 16. *Given finite Abelian groups $\mathsf{G}_0 \ldots, \mathsf{G}_r$, let $C = C_r \circ \cdots \circ C_1$ be a product cipher of independent Markov ciphers $C_i : \mathsf{G}_{i-1} \longrightarrow \mathsf{G}_i$. For any $\chi_0 \in \widehat{\mathsf{G}}_0$ and $\chi_r \in \widehat{\mathsf{G}}_r$ we have*

$$\mathrm{ELP}^C(\chi_0, \chi_r) = \textstyle\sum_{\chi_1 \in \widehat{\mathsf{G}}_1} \cdots \sum_{\chi_{r-1} \in \widehat{\mathsf{G}}_{r-1}} \prod_{i=1}^r \mathrm{ELP}^{C_i}(\chi_{i-1}, \chi_i).$$

It is a common mistake to mix up this result with the hypothesis of stochastic independence. This is a *real* equality which depends on *no* heuristic assumptions.

Proof (Sketch). Recall that a Markov cipher $C : \mathsf{G} \longrightarrow \mathsf{G}'$ between two groups G and G' is a random mapping such that for any $\delta \in \mathsf{G}$ and $\delta' \in \mathsf{G}'$ the probability $\Pr[C(x + \delta) = C(x) + \delta']$ does not depend on x.

A straightforward proof is provided in [51] for the binary case. We only have to rephrase it using characters. As a classical result (see e.g. [29]) we easily obtain

$$\mathrm{EDP}^{C}(\delta_0, \delta_r) = \sum_{\delta_1 \in \mathsf{G}_1} \cdots \sum_{\delta_{r-1} \in \mathsf{G}_{r-1}} \prod_{i=1}^{r} \mathrm{EDP}^{C_i}(\delta_{i-1}, \delta_i).$$

Then we simply apply r Fourier transforms. □

Given d plaintext/ciphertext pairs $z_i = (M_i, \mathsf{C}(M_i))$, this geometrically means that the expected value of $\mathrm{sa}(\boldsymbol{z}^d; (\overline{\chi}, \rho))$ lies on a circle of squared radius equal to $\mathrm{LP}^{\mathsf{C}}(\chi, \rho)$, its exact position on the circle depending on $\prod_j \chi_j(k_j)$.

4 A $\mathbf{Z}_{100}^{16}$ Linear Cryptanalysis of TOY100

In [13], Granboulan et al. introduce TOY100, a block cipher that encrypts blocks of 32 decimal digits. The structure of TOY100 is similar to that of the AES. An r rounds version of TOY100 is made of $r-1$ identical rounds followed by a slightly different final round. Each block is represented as a 4×4 matrix $A = (a_{i,j})_{i,j \in \{0,\ldots,3\}}$, the $a_{i,j}$'s being called subblocks. Round i (for $i = 1, \ldots, r-1$) first adds modulo 100 a subkey to each subblocks (we do not describe the key schedule here as we assume that the round keys are mutually independent), then applies a fixed substitution box to each resulting subblocks, and finally mixes the subblocks together by applying a linear transformation. The last round replaces the diffusion layer by a modulo 100 subkey addition. The round key addition, confusion and diffusion layers are respectively denoted $\sigma[K]$, γ, and θ. The diffusion layer can be represented as a matrix product $M \times A \times M$ where

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

and where all computations are performed modulo 100. The best attack against TOY100 is based on the generalization of linear cryptanalysis suggested in [13]. It breaks TOY100 reduced to 7 rounds with a data/time complexity of $0.66 \cdot 10^{31}$. We propose here a linear cryptanalysis that breaks up to 8 rounds. We first

Table 1. Complexities of the best linear cryptanalysis we obtained on reduced round versions of TOY100.

r	Lower bound on $\max_{\alpha_0,\alpha_{r-2}} \mathrm{ELP}^{(\theta\circ\gamma\circ\sigma[K])^{r-2}\circ\theta}(\alpha_0, \alpha_{r-2})$	Data/Time Complexity of the attack against r rounds
4	$0.37 \cdot 10^{-9}$	$0.27 \cdot 10^{10}$
5	$0.47 \cdot 10^{-14}$	$0.21 \cdot 10^{15}$
6	$0.66 \cdot 10^{-19}$	$0.15 \cdot 10^{20}$
7	$0.10 \cdot 10^{-23}$	$0.97 \cdot 10^{24}$
8	$0.18 \cdot 10^{-28}$	$0.55 \cdot 10^{29}$
9	$0.34 \cdot 10^{-33}$	$0.30 \cdot 10^{34}$

observe that any block

$$A(\delta) = \begin{pmatrix} \delta & 0 & 100-\delta & 0 \\ 0 & 0 & 0 & 0 \\ 100-\delta & 0 & \delta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\delta \in \{1, \ldots, 99\}$ is such that $M \times A(\delta) \times M = A(\delta)$, i.e., is not changed by the diffusion layer. We let $\mathcal{I} = \{A(\delta), \delta = 1, \ldots, 99\}$ be the set of these 99 blocks. Our attack against TOY100 reduced to r rounds first guesses 4 subblocks of the first round key and 4 subblocks of the last (the positions of which exactly correspond to the non-zero subblocks of $A(\delta)$). This allows to peel-off the first and last layers of substitution boxes, so that we now consider the transformation $(\theta \circ \gamma \circ \sigma[K])^{r-2} \circ \theta$ (where it is understood that the round keys are mutually independent). For any 4×4 input/output masks (i.e., blocks) $\alpha = (\alpha_{i,j})_{i,j \in \{1,\ldots,4\}}$ and $\beta = (\beta_{i,j})_{i,j \in \{1,\ldots,4\}}$ we let, for any transformation C on $\mathbf{Z}_{100}^{16}$,

$$\mathrm{ELP}^{\mathsf{C}}(\alpha, \beta) = \left| \mathrm{E}_M \big(\overline{\varphi_\alpha}(M) \varphi_\beta(\mathsf{C}(M)) \big) \right|^2 \quad \text{where} \quad \varphi_\alpha(M) = e^{\frac{2\pi i}{100} \sum_{i,j=1}^{4} \alpha_{i,j} m_{i,j}}.$$

Applying Theorem 16 and the observation on the diffusion layer of TOY100 we obtain that the linear probability on $(\theta \circ \gamma)^{r-2} \circ \theta$ with input (resp. output) masks $\alpha_0 \in \mathcal{I}$ (resp. $\alpha_{r-2} \in \mathcal{I}$) is such that

$$\begin{aligned} \mathrm{ELP}^{(\theta\circ\gamma\circ\sigma[K])^{r-2}\circ\theta}(\alpha_0, \alpha_{r-2}) &= \mathrm{ELP}^{(\theta\circ\gamma\circ\sigma[K])^{r-2}}(\alpha_0, \alpha_{r-2}) \\ &= \sum_{\alpha_1 \in \mathbf{Z}_{100}^4} \cdots \sum_{\alpha_{r-3} \in \mathbf{Z}_{100}^4} \prod_{i=1}^{r-2} \mathrm{ELP}^{\theta\circ\gamma\circ\sigma[K]}(\alpha_{i-1}, \alpha_i) \\ &\geq \sum_{\alpha_1 \in \mathcal{I}} \cdots \sum_{\alpha_{r-3} \in \mathcal{I}} \prod_{i=1}^{r-2} \mathrm{ELP}^{\theta\circ\gamma\circ\sigma[K]}(\alpha_{i-1}, \alpha_i) \\ &= \sum_{\alpha_1 \in \mathcal{I}} \cdots \sum_{\alpha_{r-3} \in \mathcal{I}} \prod_{i=1}^{r-2} \mathrm{LP}^{\gamma}(\alpha_{i-1}, \alpha_i). \end{aligned}$$

Practical computations of the previous equations are given in Table 1. Using an 8-round linear hull and guessing the necessary keys on an extra round, we can thus break 9 rounds of TOY100 with data complexity $0.55 \cdot 10^{29}$. We can prove that the time complexity is similar by using classical algorithmic tricks from linear cryptanalysis techniques.

5 A Generalized Cryptanalysis of SAFER K/SK

5.1 A Short Description of SAFER K/SK and Previous Cryptanalysis

The encryption procedures of SAFER K-64, SAFER K-128, SAFER SK-64, and SAFER SK-128 are almost identical. They all iterate the exact same round function, the only difference being that the recommended number of iteration of this

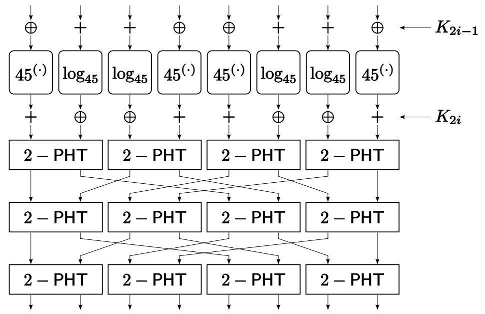

Fig. 2. The ith encryption round function of SAFER.

round function is 6 for SAFER K-64 [31], 8 for SAFER SK-64 [33], and 10 for both 128-bit versions of SAFER [31, 33]. The round function is represented on Figure 2. An r-round version of SAFER encrypts 8 bytes of text by applying the round function r times followed by a final mixed key addition (whose structure is identical to the first mixed key addition layer of the round function). Each round is parameterized by two 8-byte round keys so that a $2r+1$ round keys must be derived from the secret key.

The round function first applies a byte-wise key addition, mixing xor's and additions modulo 256. Then, each byte goes through a substitution box. Two kinds of boxes are used on SAFER: $x \mapsto (45^x \bmod 257) \bmod 256$ and its inverse. The output of the substitution box layer goes through another byte-wise key addition before being processed by a diffusion layer made of boxes called 2-PHT and defined by 2-PHT$(a, b) = (2a+b, a+b)$, the addition being performed modulo 256. Denoting $x \in \mathbf{Z}_{256}^8$ the input of the linear layer, the output $y \in \mathbf{Z}_{256}^8$ can be written as $y = M \times x$ where

$$M = \begin{pmatrix} 8 & 4 & 4 & 2 & 4 & 2 & 2 & 1 \\ 4 & 2 & 2 & 1 & 4 & 2 & 2 & 1 \\ 4 & 4 & 2 & 2 & 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 & 2 & 2 & 1 & 1 \\ 4 & 2 & 4 & 2 & 2 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 \\ 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Finally, we adopt a special notation to denote *reduced-round* versions of SAFER. We consider each of the four round layers as one fourth of a complete round. Consequently, a 2.5 reduced-round version of SAFER corresponds to two full rounds followed by the first mixed key addition and substitution layer of the third round. With these notations, the encryption procedure of SAFER K-64 is actually made of 6.25 rounds. To be consistent with the notations of the original publications, when we refer to a r-round version of SAFER, we actually mean a $r + 0.25$ reduced-round version of SAFER.

For the sake of simplicity, we restrict to give the dependencies of each round key bytes with respect to the main secret key instead of describing the key schedules of the various versions of SAFER.

Table 2. Cryptanalytic results on SAFER K/SK.

Type	# rounds	Type of the Attack	Time	Plaintexts	Reference
SAFER K	2	KPA	2^{29}	2^{13}	This paper
SAFER SK	2	KPA	2^{37}	2^{13}	This paper
SAFER K-64	3	KPA/Weak keys	2^{12}	2^{12}	[39]
SAFER K/SK	3	KPA	2^{36}	2^{36}	This paper
SAFER K-64	4	KPA/Weak keys	2^{28}	2^{28}	[39]
SAFER K/SK	4	KPA	2^{47}	2^{47}	This paper
SAFER K-64	5	KPA/Weak keys	2^{58}	2^{58}	[39]
SAFER K	5	CPA	2^{61}	2^{39}	[26, 27]
SAFER K-64	5	CPA	2^{49}	2^{44}	[26, 27]
SAFER K/SK-64	5	CPA	2^{46}	2^{38}	[52]
SAFER K/SK	5	KPA	2^{59}	2^{59}	This paper
SAFER K/SK-64	6	CPA	2^{61}	2^{53}	[52]

- SAFER K-64: The jth round key byte ($1 \leq j \leq 8$) only depends on the jth main secret key byte. For example, guessing the third byte of the main secret key allows to derive the third byte of each round key.
- SAFER SK-64: The jth byte ($1 \leq j \leq 8$) of round key number i ($1 \leq i \leq 2r+1$), depends on the ℓth byte of the secret key, where $\ell = (i+j-2) \bmod 9+1$ and where the 9th byte of the secret key is simply the xor of its previous 8 bytes.

In our analysis we assume that the key is a full vector of subkeys. When studying the average complexity of our attack, we further assume that these subkeys are randomly picked with uniform distribution.

Previous Cryptanalysis (see Table 2). Known attacks against SAFER are summarized in Table 2. The resistance of SAFER against differential cryptanalysis [5] was extensively studied by Massey in [32], where it is argued that 5 rounds are sufficient to resist to this attack. It is shown by Knudsen and Berson [26, 27] that 5 rounds can actually be broken using truncated differentials [25], a result which is extended to 6 rounds by Wu et al. in [52]. In [15], Harpes et al. apply a generalization of linear cryptanalysis [34] to SAFER K-64 but do not manage to find an effective homomorphic threefold sum for 1.5 rounds or more. Nakahara et al. showed in [39] that for certain weak key classes, one can find a 3.75-round non-homomorphic linear relation with bias $\epsilon = 2^{-29}$ (which leads to a time/plaintext complexity of $1/\epsilon^2 = 2^{58}$ known plaintexts on five rounds).

The diffusion properties of the linear layer of SAFER have also been widely studied and, compared to the confusion layer, seem to be its major weakness. In [38], Murphy proposes an algebraic analysis of the 2-PHT layer, showing in particular that by considering the message space as a **Z**-module, one can find a particular submodule which is an *invariant* of the 2-PHT transformation. In [49], Vaudenay shows that by replacing the original substitution boxes in a 4 round version of SAFER by random permutations, one obtains in 6.1% of the cases a construction that can be broken by linear cryptanalysis. This also lead Brincat and Meijer to explore potential alternatives of the 2-PHT layer [7]. The other major weakness of SAFER K is indubitably its key schedule. The analysis proposed in [26, 38] lead Massey to choose the one proposed by Knudsen in [26] for SAFER SK.

5.2 Linear Cryptanalysis of SAFER: from $\mathbf{Z}_2^8$ to $\mathbf{Z}_{2^8}$

A possible reason why linear cryptanalysis does not seem to be a threat for SAFER is that Matsui's linear characteristics (that fits so well the operations made in DES) are in fact *not* linear when it comes to the diffusion layer of SAFER except when they only focus on the least significant bit of the bytes. Yet, those bits are not biased through the substitution boxes [49]. Indeed, whereas a classical linear cryptanalysis combines text and key bits by performing xor's (i.e., additions in $\mathbf{Z}_2$), SAFER mostly relies on additions in $\mathbf{Z}_{2^8}$. In other words, the group structure that is classically assumed in linear cryptanalysis does not fit when it comes to study SAFER. We will thus focus on the additive group $(\mathbf{Z}_{256}^r, +)$. As noted already in Section 2.4, the 256^r characters of this group are called *additive character modulo* 256 and are the $\chi_{\boldsymbol{a}}$'s for $\boldsymbol{a} = (a_1, \ldots, a_r) \in [0, 255]^r$ defined by $\chi_{\boldsymbol{a}}(\boldsymbol{x}) = e^{\frac{2\pi i}{256}\sum_{\ell=1}^{r} a_\ell x_\ell}$ for all $\boldsymbol{x} = (x_1, \ldots, x_r) \in \mathbf{Z}_{256}^r$. The attack on SAFER will only involve additive characters modulo 256. To simplify the notation (and to somehow stick to the vocabulary we are used to in classical linear cryptanalysis), we denote in this section the linear probability of C with respect to $\chi_{\boldsymbol{a}}$ and $\chi_{\boldsymbol{b}}$ by $\mathrm{LP}^{\mathsf{C}}(\boldsymbol{a}, \boldsymbol{b})$ instead of $\mathrm{LP}^{\mathsf{C}}(\chi_{\boldsymbol{a}}, \chi_{\boldsymbol{b}})$. We call it the linear probability of C with input mask $\boldsymbol{a}$ and output mask $\boldsymbol{b}$.

Hiding the $\mathbf{Z}_2^8$ Group. As the encryption procedure uses additions modulo 256 together with bit-wise exclusive or, we have to deal with two types of characters. Nevertheless, one can notice that the mixture of group operations only occurs within the *confusion* layer. To simplify the analysis we can think of the succession of a round key xor and a fixed substitution box as a *keyed substitution box* (see Figure 3). Using this point of view, we represent one round of SAFER in Figure 4.

Studying SAFER's Building Blocks. Most of the building of blocks of SAFER were already considered in Section 3.2. With the notations used in this section,

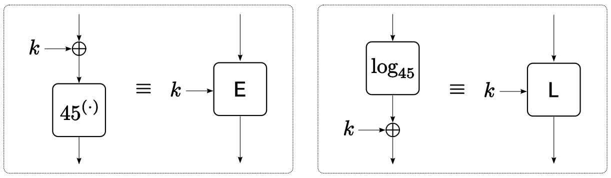

Fig. 3. Viewing key xor and fixed substitution boxes as keyed substitution boxes.

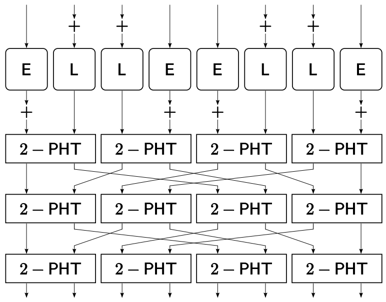

Fig. 4. Another view of SAFER.

the study of the building block (c) can be written as $\mathrm{LP}^{\cdot+k}(a, a) = 1$, where a and k are arbitrary values of $\mathbf{Z}_{256}$. If the key K is random, the previous equation implies that $\mathrm{E}_K(\mathrm{LP}^{\cdot+K}(a, a)) = 1$. Building block (b) allows to deal with the 2-PHT transformation (which is a homomorphism of $\mathbf{Z}_{256}^2$): denoting by $\boldsymbol{a} = (a_1, a_2) \in \mathbf{Z}_{256}^2$ and $\boldsymbol{b} = (b_1, b_2) \in \mathbf{Z}_{256}^2$ the input and output masks on this transformation, and noting that the 2-PHT transformation is a *symmetric* linear operator (in the sense that $\mathsf{2\text{-}PHT}^T = \mathsf{2\text{-}PHT}$), $\mathrm{LP}^{\mathsf{2\text{-}PHT}}(\boldsymbol{a}, \boldsymbol{b}) = 1 \Leftrightarrow \boldsymbol{a} = \mathsf{2\text{-}PHT}(\boldsymbol{b})$. Using the same notations, it is easy to show that when considering the parallel computation of two fixed substitution boxes S_1 and S_2 over $\mathbf{Z}_{256}$, $\mathrm{LP}^{\mathsf{S}_1\|\mathsf{S}_2}(\boldsymbol{a}, \boldsymbol{b}) = \mathrm{LP}^{\mathsf{S}_1}(a_1, b_1) \cdot \mathrm{LP}^{\mathsf{S}_2}(a_2, b_2)$. When the boxes are random and independent, this leads to $\mathrm{E}_{\mathsf{S}_1,\mathsf{S}_2}(\mathrm{LP}^{\mathsf{S}_1\|\mathsf{S}_2}(\boldsymbol{a}, \boldsymbol{b})) = \mathrm{E}_{\mathsf{S}_1}(\mathrm{LP}^{\mathsf{S}_1}(a_1, b_1)) \cdot \mathrm{E}_{\mathsf{S}_2}(\mathrm{LP}^{\mathsf{S}_2}(a_2, b_2))$.

Assuming that the key bits are mutually independent, we can now compute the linear probability of *one* full round of SAFER. Indeed if an input/output pair of masks $\boldsymbol{a} = (a_1, \ldots, a_8), \boldsymbol{b} = (b_1, \ldots, b_8)$ are given, and letting $\boldsymbol{b}' = M^T \times \boldsymbol{b} = (b'_1, \ldots, b'_8)$ (where M is the matrix defined in Section 5.1), then the linear probability on one full round, simply denoted Round, is given by

$$\mathrm{ELP}^{\mathsf{Round}}(\boldsymbol{a}, \boldsymbol{b}) = \textstyle\prod_{i=1}^{8} \mathrm{ELP}^{\mathsf{S}_i}(a_i, b'_i)$$

where S_i corresponds to a keyed E box for $i = 1, 4, 5, 8$ and to a keyed L otherwise.

5.3 Considering Several Rounds of SAFER: the Reduced Hull Effect

When several rounds are considered, Nyberg's linear hull effect [41] applies just as for classical linear cryptanalysis of Markov ciphers (see Theorem 16). Considering a succession of $r > 1$ rounds with independent round keys, and denoting $\boldsymbol{a}_0$ and $\boldsymbol{a}_r$ the input and the output masks respectively, this leads to

$$\mathrm{ELP}^{\mathsf{Round}_r \circ \cdots \circ \mathsf{Round}_1}(\boldsymbol{a}_0, \boldsymbol{a}_r) = \textstyle\sum_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{r-1}} \prod_{i=1}^{r} \mathrm{ELP}^{\mathsf{Round}_i}(\boldsymbol{a}_{i-1}, \boldsymbol{a}_i).$$

When cryptanalyzing a block cipher, it is often considered that one specific characteristic (i.e., a succession of $r + 1$ masks $\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$) is overwhelming (i.e., approximates the hull) so that

$$\mathrm{ELP}^{\mathsf{Round}_r \circ \cdots \circ \mathsf{Round}_1}(\boldsymbol{a}_0, \boldsymbol{a}_r) \approx \textstyle\prod_{i=1}^{r} \mathrm{ELP}^{\mathsf{Round}_i}(\boldsymbol{a}_{i-1}, \boldsymbol{a}_i).$$

This approach was taken by Matsui when cryptanalyzing DES. In that particular case, the correctness of this approximation could be experimentally verified [34]. In this paper we do not consider the full linear hull effect nor restrict ourselves to one specific characteristics. Instead, we consider the characteristics among the hull following a specific *pattern*.

Definition 17. *Let $\boldsymbol{a} \in \mathbf{Z}_{256}^8$ be an arbitrary mask. The* pattern *corresponding to the mask $\boldsymbol{a}$ is the binary vector of length eight, with zeroes at the zero position of $\boldsymbol{a}$ and $*$ at the non-zero positions of $\boldsymbol{a}$. The weight $w(\mathsf{p})$ of a pattern p is the number of $*$ in this pattern. We denote the fact that a mask $\boldsymbol{a}$ corresponds to pattern $\boldsymbol{p}$ by $\boldsymbol{a} \in \mathsf{p}$. We denote by* `and` *the byte-wise masking operation, i.e., given an element $m \in \mathbf{Z}_{256}^8$ and a pattern p, $m' = m \,\mathtt{and}\, \mathsf{p}$ is such that $m'_i = 0$ if $\mathsf{p}_i = 0$ and $m'_i = m_i$ otherwise, for $i = 1, \ldots, 8$. We denote by $\mathtt{int}_\mathsf{p}(m)$ the integer representation of the concatenation of the bytes of $m \,\mathtt{and}\, \mathsf{p}$ corresponding to the non-zero positions of p, and by $\mathcal{I}(\mathsf{p}) = \{\mathtt{int}_\mathsf{p}(m) : m \in \mathbf{Z}_{256}^8\}$. Finally, for an arbitrary integer $i \in \mathcal{I}(\mathsf{p})$, we denote $\mathtt{int}_\mathsf{p}^{-1}(i)$ the element $m \in \mathsf{p}$ such that $\mathtt{int}_\mathsf{p}(m) = i$.*

For example, the pattern corresponding to $\boldsymbol{a} =$ `[0,128,0,0,0,255,7,1]` is $\mathsf{p} =$ `[0*000***]` (of weight 4). If $m =$ `[3,128,128,255,0,255,7,1]`, then $m \,\mathtt{and}\, \mathsf{p} = \boldsymbol{a}$, and $\mathtt{int}_\mathsf{p}(m) = 10000000\,11111111\,00000111\,00000001_2$. Note that for an arbitrary element $m \in \mathbf{Z}_{256}^8$ and any pattern p, $\mathtt{int}_\mathsf{p}^{-1}(\mathtt{int}_\mathsf{p}(m)) = m \,\mathtt{and}\, \mathsf{p}$.

The fact that we only consider, among the hull, the characteristics following a given sequence of pattern $\mathsf{p}_0, \mathsf{p}_1, \ldots, \mathsf{p}_r$ can be written as

$$\mathrm{ELP}^{\mathsf{Round}_r \circ \cdots \circ \mathsf{Round}_1}(\boldsymbol{a}_0, \boldsymbol{a}_r) \approx \sum_{\substack{\boldsymbol{a}_1 \in \mathsf{p}_1 \\ \cdots \\ \boldsymbol{a}_{r-1} \in \mathsf{p}_{r-1}}} \prod_{i=1}^{r} \mathrm{ELP}^{\mathsf{Round}_i}(\boldsymbol{a}_{i-1}, \boldsymbol{a}_i). \tag{5}$$

where $\boldsymbol{a}_0 \in \mathsf{p}_0$ and $\boldsymbol{a}_r \in \mathsf{p}_r$. We call this approximation the *reduced hull effect*. Note that in any case, (5) actually underestimates the true linear hull.

5.4 Sketching the Construction of Reduced Hulls on Two Rounds

In order to construct such *reduced* hulls on SAFER, we start by enumerating the *possible* sequences of patterns $\mathsf{p}_1 \xrightarrow{n} \mathsf{p}_2$ on the linear diffusion layer, where n denotes the number of *distinct* pairs of input/output masks following the pattern $\mathsf{p}_1/\mathsf{p}_2$.[4] We store these sequences in tables (that we do not report here due to space constraints) that we order according to the input/output weights

[4] For example, on the linear layer, the output mask `[128,0,0,0,0,0,0,0]` corresponds to the input mask `[0,0,0,0,0,0,0,128]`. Moreover, there is no other possible mask with the same input/output patterns, which is denoted `[0000000*]` $\xrightarrow{1}$ `[*0000000]`. Two distinct pairs of masks on the linear layer following the input pattern input pattern `[0000000*]` and the output pattern `[***0*000]` can be found (namely, `[0,...,0,64]` corresponds to `[192,128,128,0,128,0,0,0]` and `[0,...,0,192]` to `[64,128,128,0,128,0,0,0]`). This is denoted `[0000000*]` $\xrightarrow{2}$ `[***0*000]`.

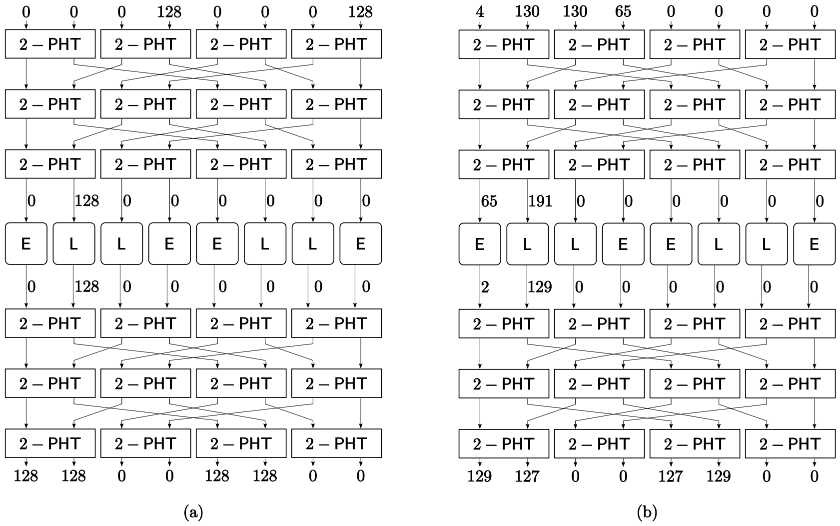

Fig. 5. The characteristics on two successive linear layers of examples 1 and 2.

$w_1 \rightarrow w_2$ $(1 \leq w_1, w_2 \leq 8)$ of the sequence $\mathsf{p}_1 \rightarrow \mathsf{p}_2$. To reduce the size of the list, we restrict it to patterns of weight sum 7 or less.

Next, we build characteristics on several rounds based on the lists of possible succession of patterns on the linear layer. We proceed step-by-step, starting with characteristics on two rounds. Two characteristics *on full rounds* can only be concatenated if the output mask of the first one is equal to the input mask of the second one. This translates for patterns as follows: two successions of patterns on the linear layer can only be concatenated if the output pattern of the first succession is equal to the input pattern of the second succession.

Example 1. We can concatenate $\texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]}$ and $\texttt{[0*000000]} \xrightarrow{1} \texttt{[**00**00]}$. We denote this by $\texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]} \xrightarrow{1} \texttt{[**00**00]}$. This means that succession of patterns of weights $2 \rightarrow 1 \rightarrow 4$ on two rounds exist. In this particular example, there is only one characteristic corresponding to this succession of masks, which is represented on Figure 5(a).

Example 2. Similarly, we can obtain the succession $\texttt{[****0000]} \xrightarrow{252} \texttt{[**000000]} \xrightarrow{254} \texttt{[**00**00]}$ which is a succession of pattern of weights $4 \rightarrow 2 \rightarrow 4$ on two rounds. In this case, $252 \times 254 = 64008$ distinct characteristics correspond to this succession (one of which is represented on Figure 5(b)).

Finally, it should be noted that the characteristic of Example 1 actually leads to an ELP equal to 0, as both input and output masks on the substitution box are equal to 128, which is equivalent to computing the traditional linear probability by only considering the least significant bit. In the second example, the computation of the reduced hull leads to a non-zero linear probability.

Table 3. Key Recovery Attack against a r reduced-round version of SAFER.

Input: A reduced hull on r rounds with input mask $\boldsymbol{a_0} \in \mathsf{p}_0$ and output mask $\boldsymbol{a_r} \in \mathsf{p}_r$.
Output: A set of counters $\mathrm{lp}_{\kappa_1,\kappa_2,\kappa_{2r+1}}$ with $\kappa_1, \kappa_2 = 0, \ldots, 2^{8w(\mathsf{p}_0)} - 1$ and $\kappa_{2r+1} = 0, \ldots, 2^{8w(\mathsf{p}_r)} - 1$.
Memory: A set of counters $N_{i,j}$ initialized to 0, with $i = 0, \ldots, 2^{8 \cdot w(\mathsf{p}_0)} - 1$ and $j = 0, \ldots, 2^{8 \cdot w(\mathsf{p}_r)} - 1$.

0: **foreach** of the d plaintext/ciphertext pair (m, c) **do**
1: $i \leftarrow \mathtt{int}_{\mathsf{p}_0}(m)$ and $j \leftarrow \mathtt{int}_{\mathsf{p}_r}(c)$
2: $N_{i,j} \leftarrow N_{i,j} + 1$
3: **done**
4: **foreach** $(\kappa_1, \kappa_2, \kappa_{2r+1}) \in \mathcal{I}(\mathsf{p}_0) \times \mathcal{I}(\mathsf{p}_0) \times \mathcal{I}(\mathsf{p}_r)$ **do**
5: $k_1 \leftarrow \mathtt{int}^{-1}_{\mathsf{p}_0}(\kappa_1)$, $k_2 \leftarrow \mathtt{int}^{-1}_{\mathsf{p}_0}(\kappa_2)$, and $k_{2r+1} \leftarrow \mathtt{int}^{-1}_{\mathsf{p}_r}(\kappa_{2r+1})$
6: `/* compute the lp corresponding to the round keys guess */`
7: $\mathcal{L} \leftarrow 0$
8: **foreach** $(i, j) \in [0, \ldots, 2^{8w(\mathsf{p}_0)} - 1] \times [0, \ldots, 2^{8w(\mathsf{p}_r)} - 1]$ such that $N_{i,j} > 0$ **do**
9: $m \leftarrow \mathtt{int}^{-1}_{\mathsf{p}_0}(i)$ and $c \leftarrow \mathtt{int}^{-1}_{\mathsf{p}_r}(j)$
10: Add/xor k_1 to m, apply the subst. box layer, add/xor k_2, call the result m'.
11: Subtract k_{2r+1} to c, call the result c'
12: $\mathcal{L} \leftarrow \mathcal{L} + N_{i,j} \cdot \overline{\chi_{\boldsymbol{a_0}}}(m') \chi_{\boldsymbol{a_r}}(c')$
13: **done**
14: $\mathrm{lp}_{\kappa_1,\kappa_2,\kappa_{2r+1}} \leftarrow |\mathcal{L}|^2$.
15: **done**

5.5 Attacks on Reduced-Round Versions of SAFER

From Distinguishing Attacks to Key Recovery. In this section, a *reduced hull on r diffusion layers of* SAFER corresponds to a succession patterns on r successive linear layers separated by confusion layers. The *weight* of a reduced hull is the number of active substitution boxes (i.e., the number of boxes with non-zero input/output masks) for any characteristic of the hull. For example, the succession $\texttt{[****0000]} \xrightarrow{252} \texttt{[**000000]} \xrightarrow{254} \texttt{[**00**00]}$ (of Example 2) is a reduced hull of weight 2 on two diffusion layers. A reduced hull easily leads to a distinguishing attack on a reduced-round version of SAFER that would start and end by a diffusion layer.

Table 3 describes a key recovery attack on a SAFER reduced to r rounds by use of a reduced hull on r diffusion layers. Each of the counters obtained with this algorithm measures the probability that the corresponding subset of round key bits (for round keys 1,2, and $2r+1$) is the correct one. We expect the correct guess to be near the top of a list sorted according to these counters when the number of plaintexts/ciphertext pairs is close to $d = 1/\mathrm{ELP}^{\mathsf{C}}(\boldsymbol{a}_0, \boldsymbol{a}_r)$.

In the worst case, line 4 loops $2^{8\cdot(2w(\mathsf{p}_0)+w(\mathsf{p}_r))}$ times. In practice, the complexity is much lower (by considering key dependence due to the key schedule) and depends on the number of bits n_k that we need to guess in our attacks. When considering SAFER K-64 for example, a guess for the meaningful bytes of k_1 uniquely determines the bytes of k_2 (for the reasons given in Section 5.1). Similarly, the meaningful bytes of k_{2r+1} that are at the same positions than those of k_1 are also uniquely determined. When considering SAFER SK-64, similar techniques may apply, depending on the specific shapes of the input/output masks and the number of rounds. In all cases, if the meaningful bytes of k_2 and k_{2r+1} are actually added modulo 256, then they don't need to be guessed (as they don't alter the linear probability). If we only consider SAFER SK, this observation also applies

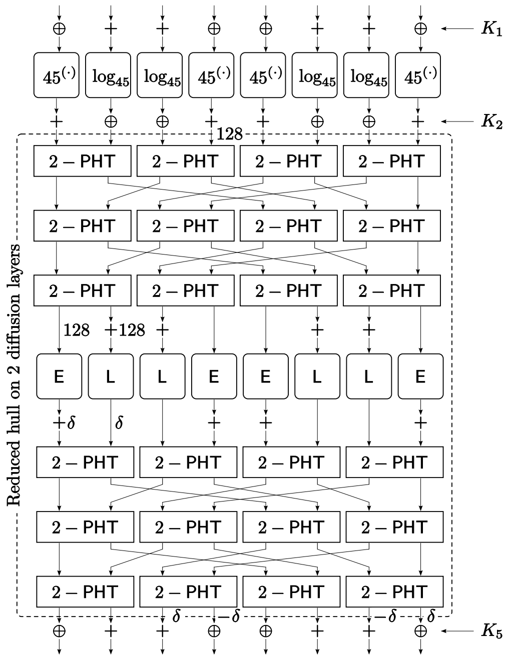

Fig. 6. Reduced hull on two diffusion layers used to attack two rounds of SAFER K.

to k_2. Finally, line 8 loops 2^{n_p} times where $n_p = \min(8 \cdot (w(\mathsf{p}_0) + w(\mathsf{p}_r)), \log_2 d)$ (as $\sum_{i,j} N_{i,j} = d$). Consequently, given any input/output masks $\boldsymbol{a}_0 \in \mathsf{p}_0$ and $\boldsymbol{a}_r \in \mathsf{p}_r$, the time complexity of the attack is given by

$$T = \frac{1}{\mathrm{ELP}^{\mathsf{C}}(\boldsymbol{a}_0, \boldsymbol{a}_r)} + 2^{n_k + n_p}. \tag{6}$$

An attack on 2 Rounds. The best attacks we could find on two rounds are based on reduced hull of weight 2 and are listed in Table 4. The best attack on SAFER K exploits the reduced hull represented on Figure 6. To perform the attack, one needs to guess 8 bits of K_1, no bits of K_2 (as those that could be meaningful are added modulo 256 and thus do not influence the linear probability), and 8 bits of K_5 (as those in position 4 are uniquely determined by the guess made on K_1). We thus obtain $n_k = 24$. The algorithm then loops through the $d = 1/\max_{\boldsymbol{a}_0, \boldsymbol{a}_2} \left(\mathrm{ELP}^{\mathsf{H}^{(2)}}(\boldsymbol{a}_0, \boldsymbol{a}_2)\right)$ pairs, where $\mathsf{H}^{(2)}$ here denotes the reduced hull and where $\boldsymbol{a}_0$ (resp. $\boldsymbol{a}_2$) denote the input (resp. output) mask on $\mathsf{H}^{(2)}$. The final complexity is computed according to (6) and given in Table 4.

For SAFER SK, the previous reduced hull leads to a higher complexity as 8 more bits of K_5 must be guessed. It appears that the best attack on two rounds of SAFER SK makes use of the first characteristics given in Table 4.

Attacks on 3, 4, and 5 Rounds. To attack three rounds of SAFER K/SK, we make use of reduced hulls on two diffusion layers of weight 6. We restricted

Table 4. Selected reduced hulls on r diffusion layers and attack complexities against r rounds of SAFER K/SK (with $\mathrm{ELP}_{\max}$ denoting $\max_{\boldsymbol{a}_0,\boldsymbol{a}_r} \mathrm{ELP}^{\mathsf{H}^{(r)}}(\boldsymbol{a}_0, \boldsymbol{a}_r)$).

r	Reduced hull	$\mathrm{ELP}_{\max}$	n_p	n_k	Complexity
2	$[000{*}0000] \xrightarrow{1} [{*}{*}000000] \xrightarrow{254} [{*}{*}00{*}{*}00]$	2^{-13}	13	24/24	$2^{37}/2^{37}$
2	$[000{*}0000] \xrightarrow{1} [{*}{*}000000] \xrightarrow{255} [00{*}{*}00{*}{*}]$	2^{-13}	13	16/24	$2^{29}/2^{37}$
3	$[0{*}000000] \xrightarrow{1} [{*}{*}00{*}{*}00] \xrightarrow{255} [000{*}000{*}] \xrightarrow{1} [0{*}000000]$	2^{-36}	16	8/16	$2^{36}/2^{36}$
4	$[000{*}0000] \xrightarrow{1} [{*}{*}000000] \xrightarrow{254} [{*}{*}00{*}{*}00] \xrightarrow{255} [000{*}000{*}] \xrightarrow{1} [0{*}000000]$	2^{-47}	16	16/24	$2^{47}/2^{47}$
5	$[000{*}0000] \xrightarrow{1} [{*}{*}000000] \xrightarrow{254} [{*}{*}00{*}{*}00] \xrightarrow{254} [0{*}000{*}00] \xrightarrow{1} [0{*}000{*}00] \xrightarrow{254} [0{*}0{*}0{*}0{*}]$	2^{-59}	40	16/24	$2^{59}/2^{59}$

our search to input/output patterns of weight 1 to limit the number of key bits guess. Using similar techniques as for the two rounds case, we manage to mount an attack against both versions of SAFER reduced to three rounds within a complexity of 2^{37} (see Table 4).

To attack four rounds, we use the reduced hull on four diffusion layers listed in Table 4. It appears that SAFER K/SK reduced to four rounds can be attacked within a complexity of 2^{47}. Whereas our generalization of linear cryptanalysis seems necessary to derive this reduced hull on four rounds, the attack itself (which only involves the input and output masks, not the intermediate ones) actually *exactly* corresponds to the original version of linear cryptanalysis: as the non-zero bytes of both input/output masks maximizing the expected linear probability are equal to 128, they only focus on one single bit. The last reduced hull of Table 4 shows that 5 rounds of SAFER K can be broken within a complexity of 2^{59}. Finally, we noted that among the output masks that maximize the expected linear probability, several end by an even byte. For example the best reduced hull is obtained when the last output masks ends by a 2. This remarks applies to the fourth byte of the output mask. Consequently, strictly less than 16 key bits need to be guessed in the last round key, so that the same reduced hull can also be used break 5 rounds of SAFER SK.

6 DEAN: a Toy Example

We introduce DEAN18 (as for Digital Encryption Algorithm for Numbers) a toy cipher that encrypts blocks of 18 decimal digits (which approximatively corresponds to a block size of 60 bits). This could be used to encrypt a credit-card number for example. The structure of the toy cipher we suggest is inspired from that of the AES [9]. We consider an R-round substitution-permutation network, each round being based on the same structure. Blocks are represented as 3×3 arrays of elements of the additive group $\mathbf{Z}_{10} \times \mathbf{Z}_{10}$. Each round successively applies to the plaintext the following operations:

- `AddRoundKeys`, that performs a digit-wise addition of a round key to the input (the addition being taken modulo 10),
- `SubBytes`, that applies a fixed bijective substitution box S (defined in Table 5, where an element $(a, b) \in \mathbf{Z}_{10}^2$ are represented as an integer $10 \cdot a + b \in [0, 99]$) on each 2-digit element of the array,

- `ShiftRows`, that shift to the left each row of the input over a given offset (equal to the row number, starting from 0 at the top),
- `MixColumns`, that multiplies each column of the input by the matrix

$$\mathsf{M} = \begin{pmatrix} \alpha & 1 & 1 \\ 1 & \alpha & 1 \\ 1 & 1 & \alpha \end{pmatrix}$$

where the multiplication of an arbitrary element $(a, b) \in \mathbf{Z}_{10}^2$ by α (resp. 1) is defined by $\alpha \cdot (a, b) = (a + b, -a)$ (resp. $1 \cdot (a, b) = (a, b)$).[5] One can easily see that this defines a structure on $\mathbf{Z}_{10}^2$ or $\mathbf{Z}_{10}^3$ that is isomorphic to $\mathrm{GF}(4) \times \mathrm{GF}(25)$ or $\mathrm{GF}(8) \times \mathrm{GF}(125)$ on which the matrix is an MDS matrix [22, 49].

The branch number of the matrix multiplication is 4, i.e., the total number of non-zero elements of the input and output columns is either 0 or 4 or more. Consequently, given a non-trivial character $\rho = (\rho_1, \rho_2, \rho_3)$ on the output of the transformation we obtain (given that we are considering a building block of type (b)) that the appropriate character $\chi = (\chi_1, \chi_2, \chi_3)$ on the input is non-trivial and that among the 6 characters $\chi_1, \ldots, \rho_3$, at least 4 are non-trivial. When at least one of the six characters is non-trivial, we say that the column is active.

Extending this result to the whole `MixColumns` transformation and applying similar arguments than those used on the AES [9], one can obtain that any two rounds characteristic (i.e., succession of three characters on the text space) has a weight lower bounded by $4Q$, where the weight is simply the number of non-trivial characters on $\mathbf{Z}_{10}^2$ among the 27 components of the three characters and Q is the number of active columns at the output of the first round. Similar arguments also lead to the fact that the sum of the number of active columns at the output of the first and of the third round of a 4-round characteristic is at least 4. Consequently, the weight of a 4-round characteristic is at least 16.

Denoting by $\mathrm{LP}_{\max}^{\mathsf{S}}$ the maximum value of $\mathrm{LP}^{\mathsf{S}}(\chi, \rho)$ over pairs of non-trivial characters, we conclude (under standard heuristic assumptions on the independence of the output of the characters at each round) that the linear probability of a $4r$-rounds characteristic is upper-bounded by $(\mathrm{LP}_{\max}^{\mathsf{S}})^{16r}$. Assuming that one characteristic among the linear hull [41] is overwhelming and that the bound given by Theorem 7 is tight, this suggest that in the best case (from an adversary point of view), a distinguishing attack against a $4r$-round version of our toy cipher needs at least $d \approx (\mathrm{LP}_{\max}^{\mathsf{S}})^{-16r}$ samples. For the substitution box of our toy cipher, we obtain $\mathrm{LP}_{\max}^{\mathsf{S}} \approx 0.069$, so that the number of samples that is necessary to attack four rounds with linear cryptanalysis is close to $3.8 \times 10^{18} \approx 2^{61}$. We conclude that $R = 8$ rounds are enough for DEAN18 to keep a high security margin (as far as linear cryptanalysis is concerned).

TOY100 [13] is a similar construction using 11 rounds and blocks of 32 digits, but where a block is a 4×4 array of $\mathbf{Z}_{100}$ elements. One problem with the algebraic structure of $\mathbf{Z}_{100}$ is that its 2-Sylow subgroup is cyclic so there are no MDS matrices. This is not the case of $\mathbf{Z}_{10}^2$.

[5] Considering the elements of $\mathbf{Z}_{10}^2$ as elements of $\mathbf{Z}_{10}[\alpha]/(\alpha^2 - \alpha + 1)$ naturally leads to this definition. One could also try to encrypt blocks of 27 digits by using $\mathbf{Z}_{10}^3$ considered as $\mathbf{Z}_{10}[\alpha]/(\alpha^3 - \alpha^2 - 1)$.

Table 5. A fixed substitution box on $\mathbf{Z}_{10}^2$.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
27	48	46	31	63	30	91	56	47	26	10	34	8	23	78	77	80	65	71	43
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
36	72	29	79	83	7	58	95	69	74	67	35	32	59	82	14	75	99	24	87
40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
16	90	76	51	28	93	50	38	25	3	13	97	55	60	49	86	57	89	62	45
60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
18	37	1	6	98	68	39	17	19	20	64	44	33	40	96	2	12	41	52	85
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
81	5	0	15	54	88	92	21	84	22	53	11	4	94	42	66	70	9	61	73

7 Conclusion

The theory developed in this paper makes it possible to generalize linear cryptanalysis to random sources and random permutations defined over sets of any cardinality. This generalization appears to be very natural as it encompasses the original one in the binary case, always preserves cumulative effects of linear hulls, and keeps the intrinsic link with differential cryptanalysis. We also showed that there always exists a group law allowing to express the best distinguisher in a linear way (yet, finding this law certainly is a hard task in general). The theory proves to be useful not only in the non binary case but also in the binary case, e.g. when a mixture of group laws is used in the block cipher design.

Acknowledgments. The authors would like to thank anonymous referees for their valuable comments, Jean Monnerat for a quite useful pointer to [40] and Matthieu Finiasz for a priceless help regarding the attack on SAFER.

References

1. C. Adams, H.M. Heys, S.E. Tavares, and M. Wiener. CAST256: a submission for the advanced encryption standard, 1998. First AES Candidate Conference (AES1).
2. T. Baignères and M. Finiasz. Dial C for Cipher. In *Selected Areas in Cryptography 2006*, LNCS. Springer-Verlag, 2006. To appear.
3. T. Baignères, P. Junod, and S. Vaudenay. How far can we go beyond linear cryptanalysis? In *Advances in Cryptology* - ASIACRYPT *'04*, volume 3329 of *LNCS*, pages 432–450. Springer-Verlag, 2004.
4. E. Biham. On Matsui's linear cryptanalysis. In [10], pages 341–355.
5. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4:3–72, 1991.
6. A. Biryukov, C. De Cannière, and M. Quisquater. On multiple linear approximations. In *Advances in Cryptology* - CRYPTO *'04*, volume 3152 of *LNCS*, pages 1–22. Springer-Verlag, 2004.
7. K. Brincat and H. Meijer. On the SAFER cryptosystem. In *Cryptography and Coding, 6th IMA International Conference*, volume 1355 of *LNCS*, pages 59–68, 1997.

8. F. Chabaud and S. Vaudenay. Links between differential and linear cryptanalysis. In [10], pages 356–365.
9. J. Daemen and V. Rijmen. AES proposal: Rijndael. NIST AES Proposal, 1998.
10. *Advances in Cryptology* - EUROCRYPT *'94*, volume 950 of *LNCS*. Springer-Verlag, 1995.
11. W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2 of *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, second edition, 1971.
12. *Fast Software Encryption 1996*, volume 1039 of *LNCS*. Springer-Verlag, 1996.
13. L. Granboulan, E. Levieil, and G. Piret. Pseudorandom permutation families over Abelian groups. In *Fast Software Encryption 2006*, volume 4047 of *LNCS*, pages 57–77. Springer-Verlag, 2006.
14. *Selected Areas in Cryptography 2004*, volume 3357 of *LNCS*. Springer-Verlag, 2004.
15. C. Harpes, G.G. Kramer, and J. Massey. A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma. In *Advances in Cryptology* - EUROCRYPT *'95*, volume 921 of *LNCS*, pages 24–38. Springer-Verlag, 1995.
16. C. Harpes and J. Massey. Partitioning cryptanalysis. In *Fast Software Encryption 1997*, volume 1267 of *LNCS*, pages 13–27. Springer-Verlag, 1997.
17. T. Jakobsen. *Higher-order cryptanalysis of block ciphers*. PhD thesis, Department of Mathematics, Technical University of Denmark, 1999.
18. T. Jakobsen and C. Harpes. Non-uniformity measures for generalized linear cryptanalysis and partitioning cryptanalysis. In PRAGOCRYPT *'96*. CTU Publishing House, 1996.
19. P. Junod. On the optimality of linear, differential and sequential distinguishers. In *Advances in Cryptology* - EUROCRYPT *'03*, volume 2656 of *LNCS*, pages 17–32. Springer-Verlag, 2003.
20. P. Junod and S. Vaudenay. Optimal key ranking procedures in a statistical cryptanalysis. In *Fast Software Encryption 2003*, volume 2887 of *LNCS*, pages 235–246. Springer-Verlag, 2003.
21. P. Junod and S. Vaudenay. FOX: a new family of block ciphers. In [14], pages 114–129.
22. P. Junod and S. Vaudenay. Perfect diffusion primitives for block ciphers. In [14], pages 84–99.
23. B. Kaliski and M. Robshaw. Linear cryptanalysis using multiple approximations. In *Advances in Cryptology* - CRYPTO *'94*, volume 839 of *LNCS*, pages 26–39. Springer-Verlag, 1994.
24. J. Kelsey, B. Schneier, and D. Wagner. mod$n$ cryptanalysis, with applications against RC5P and M6. In *Fast Software Encryption 1999*, volume 1636 of *LNCS*, pages 139–155. Springer-Verlag, 1999.
25. L. Knudsen. Truncated and higher order differentials. In [43], pages 196–211.
26. L. Knudsen. A detailed analysis of SAFER K. *Journal of Cryptology*, 13(4):417–436, 2000.
27. L. Knudsen and T. Berson. Truncated differentials of SAFER. In [12], pages 15–26.
28. L. Knudsen and M. Robshaw. Non-linear approximations in linear cryptanalysis. In *Advances in Cryptology* - EUROCRYPT *'96*, volume 1070 of *LNCS*, pages 224–236. Springer-Verlag, 1996.
29. X. Lai, J. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. In *Advances in Cryptology* - EUROCRYPT *'91*, volume 547 of *LNCS*, pages 17–38. Springer-Verlag, 1991.

30. X. Lai and J.L. Massey. A proposal for a new block encryption standard. In *Advances in Cryptology* - EUROCRYPT *'90*, volume 473 of *LNCS*, pages 389–404. Springer-Verlag, 1991.
31. J. Massey. SAFER-K64: a byte-oriented block-ciphering algorithm. In *Fast Software Encryption 1993*, volume 809 of *LNCS*, pages 1–17. Springer-Verlag, 1994.
32. J. Massey. SAFER-K64: one year later. In [43], pages 212–241.
33. J. Massey. Strengthened key schedule for the cipher SAFER. Posted on USENET newsgroup sci.crypt, September 9, 1995.
34. M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology* - CRYPTO *'94*, volume 839 of *LNCS*, pages 1–11. Springer-Verlag, 1994.
35. M. Matsui. New structure of block ciphers with provable security against differential and linear cryptanalysis. In [12], pages 205–218.
36. A. Menezes, P. Van Oorschot, and S. Vanstone. *Handbook of applied cryptography.* The CRC Press series on discrete mathematics and its applications. CRC-Press, 1997.
37. M. Minier and H. Gilbert. Stochastic cryptanalysis of Crypton. In *Fast Software Encryption 2000*, volume 1978 of *LNCS*, pages 121–133. Springer-Verlag, 2001.
38. S. Murphy. An analysis of SAFER. *Journal of Cryptology*, 11(4):235–251, 1998.
39. J. Nakahara, B. Preneel, and J. Vandewalle. Linear cryptanalysis of reduced-round versions of the SAFER block cipher family. In *Fast Software Encryption 2000*, volume 1978 of *LNCS*, pages 244–261. Springer-Verlag, 2001.
40. Melvyn B. Nathanson. *Elementary Methods in Number Theory.* Graduate Texts in Mathematics. Springer-Verlag, 2000.
41. K. Nyberg. Linear approximation of block ciphers. In [10], pages 439–444.
42. M. Parker. Generalized S-Box linearity. Technical report, NESSIE Project, 2003. `https://www.cryptonessie.org`.
43. *Fast Software Encryption 1994*, volume 1008 of *LNCS*. Springer-Verlag, 1995.
44. J.A. Rice. *Mathematical Statistics and Data Analysis.* Duxbury Press, 2nd edition, 1995.
45. Rich Schroeppel. Hasty pudding cipher specification, June 1998. Available on `http://www.cs.arizona.edu/~rcs/hpc/hpc-spec`.
46. T. Shimoyama and T. Kaneko. Quadratic relation of S-Box and its application to the linear attack of full round DES. In *Advances in Cryptology* - CRYPTO *'98*, volume 1462 of *LNCS*, pages 200–211. Springer-Verlag, 1998.
47. F.-X. Standaert, G. Rouvroy, G. Piret, J.-J. Quisquater, and J.-D. Legat. Key-dependent approximations in cryptanalysis: an application of multiple Z4 and non-linear approximations. In *24th Symposium on Information Theory in the Benelux*, 2003.
48. J. Stern and S. Vaudenay. CS-Cipher. In *Fast Software Encryption 1998*, volume 1372 of *LNCS*, pages 189–204. Springer-Verlag, 1998.
49. S. Vaudenay. On the need for multipermutations: Cryptanalysis of MD4 and SAFER. In [43], pages 286–297.
50. S. Vaudenay. An experiment on DES statistical cryptanalysis. In *3rd ACM Conference on Computer and Communications Security*, pages 139–147. ACM Press, 1996.
51. S. Vaudenay. On the security of CS-cipher. In *Fast Software Encryption 1999*, volume 1636 of *LNCS*, pages 260–274. Springer-Verlag, 1999.
52. H. Wu, F. Bao, R. Deng, and Q.-Z. Ye. Improved truncated differential attacks on SAFER. In *Advances in Cryptology* - ASIACRYPT *'98*, volume 1514 of *LNCS*, pages 133–147. Springer-Verlag, 1998.