Towards a unifying view of block cipher cryptanalysis

David Wagner[⋆]

University of California, Berkeley

Abstract. We introduce *commutative diagram cryptanalysis*, a framework for expressing certain kinds of attacks on product ciphers. We show that many familiar attacks, including linear cryptanalysis, differential cryptanalysis, differential-linear cryptanalysis, mod n attacks, truncated differential cryptanalysis, impossible differential cryptanalysis, higher-order differential cryptanalysis, and interpolation attacks can be expressed within this framework. Thus, we show that commutative diagram attacks provide a unifying view into the field of block cipher cryptanalysis. Then, we use the language of commutative diagram cryptanalysis to compare the power of many previously known attacks. Finally, we introduce two new attacks, *generalized truncated differential cryptanalysis* and *bivariate interpolation*, and we show how these new techniques generalize and unify many previous attack methods.

1 Introduction

How do we tell if a block cipher is secure? How do we design good ciphers? These two questions are central to the study of block ciphers, and yet, after decades of research, definitive answers remain elusive. For the moment, the art of cipher evaluation boils down to two key tasks: we strive to identify as many novel cryptanalytic attacks on block ciphers as we can, and we evaluate new designs by how well they resist known attacks.

The research community has been very successful at this task. We have accumulated a large variety of different attack techniques: differential cryptanalysis, linear cryptanalysis, differential-linear attacks, truncated differential cryptanalysis, higher-order differentials, impossible differentials, mod n attacks, integrals, boomerangs, sliding, interpolation, the yo-yo game, and so on. The list continues to grow. Yet, how do we make sense of this list? Are there any common threads tying these different attacks together?

In this paper, we seek unifying themes that can put these attacks on a common foundation. We have by no means accomplished such an ambitious goal; rather, this paper is intended as a first step in that direction. In this paper, we show how a small set of ideas can be used to generate many of today's known attacks. Then, we show how this viewpoint allows us to compare the strength of different types of attacks, and possibly to discover new attack techniques. We

[⋆] This work is supposed by NSF ITR CNS-0113941.

$$\begin{tikzcd}
A \arrow[r, "g"] \arrow[d, "f"'] & X \arrow[d, "i"] \\
B \arrow[r, "h"] & Y
\end{tikzcd}$$

Fig. 1. An example of a commutative diagram. The intended meaning of this diagram is that $h \circ f = i \circ g$.

hope this perspective will be of some interest, if only to see a different way to think about the known cryptanalytic attacks on block ciphers.

2 Background

What is a block cipher? A block cipher is a map $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ so that E_k is invertible for all keys $k \in \mathcal{K}$, and both E_k and E_k^{-1} can be efficiently computed. The set $\mathcal{M}$ is the space of texts; for instance, for AES, it is $\mathcal{M} = \{0, 1\}^{128}$.

When is a block cipher secure? A block cipher is secure if it behaves as a pseudorandom permutation. In other words, it must be secure against *distinguishing attacks*: no efficient algorithm A given interactive access to encryption and decryption black boxes should be to distinguish the real cipher (i.e., E_k and E_k^{-1}) from a truly random permutation (i.e., π and π^{-1}, where π is uniformly distributed on the set of all permutations on $\mathcal{M}$) with non-negligible advantage. The distinguishing advantage of an attack A is given by $\mathrm{Adv}\, A = \Pr[A^{E_k, E_k^{-1}} = 1] - \Pr[A^{\pi, \pi^{-1}} = 1]$.

In this paper, we focus exclusively on distinguishing attacks. Usually, once a distinguishing attack is found, a key-recovery attack soon follows; the hard part is in finding a distinguishing attack in the first place, or in building a cipher secure against distinguishing attacks.

How are block ciphers built? Most block ciphers are *product ciphers*. In other words, the cipher is built as the composition of individual round transformations: we choose a round function $f : \mathcal{M} \to \mathcal{M}$, compute a sequence of round keys $k_1, \ldots, k_n$ as a function of the key k, and set $E_k = f_{k_n} \circ \cdots \circ f_{k_1}$. The function f computes one round of the cipher.

Commutative diagrams. In the discussion to follow, it will be useful to introduce a concept from abstract algebra: that of *commutative diagrams*. Commutative diagrams are a concise notation for expressing functional composition properties. An example of a commutative diagram can be found in Fig. 1. In this example, the symbols A, B, X, Y represent sets, and the symbols f, g, h, i are functions with signatures $f : A \to B$, $g : A \to X$, $h : B \to Y$, and $i : X \to Y$. We say that the diagram "commutes" if $h \circ f = i \circ g$, or in other words, if $h(f(a)) = i(g(a))$ for all $a \in A$. Notice how paths correspond to functions, obtained by composing the maps associated with each edge in the path. In this diagram, there are two paths from A to Y, corresponding to two functions with signature $A \to Y$. Informally, the idea is that it doesn't matter which path we follow from A to Y; we will obtain the same map either way. More complicated diagrams can be used to express more complex relationships, and identifying the set of implied identities is merely a matter of chasing arrows through the diagram.

Markov processes. Also, we recall the notion of Markov processes. A Markov process is a pair of random variables I, J, and to it we associate a transition matrix M given by $M_{i,j} = \Pr[J = j | I = i]$. We call the sequence of random variables $I - J - K$ a Markov chain if K is conditionally independent of I given J. We can associate transition matrices M, M', M'' to the Markov processes $I - J$, $J - K$, and $I - K$, respectively, and if $I - J - K$ forms a Markov chain, we will have $M'' = M' \cdot M$. In other words, composition of Markov processes corresponds to multiplication of their associated transition matrices.

The maximum advantage of an adversary at distinguishing one Markov process from another can be calculated using decorrelation theory [14]. Let $||M||_\infty$ denote the ℓ_∞ norm of the matrix M, i.e., $||M||_\infty = \max_i \sum_j |M_{i,j}|$. We can consider an adversary A who is allowed to choose a single input, feed it through the Markov process, and observe the corresponding output. The maximum advantage of any such adversary at distinguishing M from M' will then be exactly $\frac{1}{2}||M - M'||_\infty$. If U denotes the uniform $m \times n$ transition matrix, i.e., $U_{i,j} = 1/n$ for all i, j, then $||M_1 M_2 - U||_\infty \leq ||M_1 - U||_\infty \cdot ||M_2 - U||_\infty$.

The above calculations can be extended to calculate the advantage $\mathrm{Adv}\, A = \Pr[A^M = 1] - \Pr[A^{M'} = 1]$ of an adversary A who can interact repeatedly with the Markov process. First, if $\frac{1}{2}||M - M'||_\infty = \epsilon$ denotes the advantage of a single-query adversary, then an adversary making q queries has advantage at most $q \cdot \epsilon$. In practice, when ϵ is small, the advantage of a q-query adversary often scales roughly as $\sim \sqrt{q} \cdot \epsilon$. Hence, as a rough rule of thumb, $\Theta(1/\epsilon^2)$ queries often are necessary and sufficient to distinguish M from M' with non-trivial probability [3]. We emphasize, though, that this heuristic is not always valid; there are many important exceptions.

The advantage of a q-query adversary can be computed more precisely. If M is a $m \times n$ matrix, let $[M]^q$ denote the $m^q \times n^q$ matrix given by $([M]^q)_{i,j} = M_{i_1,j_1} \times \cdots \times M_{i_q,j_q}$. Define the matrix norm $||M||_a = \max_{i_1} \sum_{j_1} \cdots \max_{i_q} \sum_{j_q} |M_{i,j}|$. Then, in an adaptive attack, the maximum advantage of any q-query adversary is exactly $\max_A \mathrm{Adv}\, A = \frac{1}{2}||[M]^q - [M']^q||_a$. In a non-adaptive attack, the maximum advantage of any q-query non-adaptive adversary is exactly $\frac{1}{2}||[M]^q - [M']^q||_\infty$.

Organization. The rest of this paper studies cryptanalysis of product ciphers. First, we describe commutative diagrams and their relevance to cryptanalysis. Then, we explore statistical attacks, a probabilistic generalization of commutative diagram attacks, and then we further generalize by introducing the notion of higher-order attacks. Finally, we explore algebraic attacks.

$$\begin{tikzcd}
\mathcal{M} \arrow[r, "\rho"] \arrow[d, "f_k"'] & Y \arrow[d, "g"] \\
\mathcal{M} \arrow[r, "\rho'"] & Y'
\end{tikzcd}$$

Fig. 2. A local property of the round function f_k.

3 Commutative diagram attacks

The basic recipe for analyzing a product cipher is simple:

1. Identify local properties of the cipher's round functions.
2. Piece these together to obtain a global property of the cipher as a whole.

In this way, we seek to exploit the structure of a product cipher—namely, its construction as a composition of round functions—to simplify the cryptanalyst's task.

How do we identify local properties of the round function that are both non-trivial and can be spliced together suitably? This is where commutative diagrams can help. Let $f_k : \mathcal{M} \to \mathcal{M}$ denote a round function. Suppose we can find a property of the input that is preserved by the round function; then this would suffice as a local property. If there is some partial information about x that allows to predict part of the value of $f_k(x)$, this indicates a pattern of non-randomness in the round function that might be exploitable.

One way to formalize this is using *projections.* A projection is a function $\rho : \mathcal{M} \to Y$ from the text space to a smaller set Y. If we have two projections ρ, ρ' so that $\rho'(f_k(x))$ can be predicted from $\rho(x)$, then we have a local property of the round function. To make this more precise, we look for projections $\rho : \mathcal{M} \to Y$, $\rho' : \mathcal{M} \to Y'$ and a function $g : Y \to Y'$ so that $\rho' \circ f_k = g \circ \rho$ for all $k \in \mathcal{K}$, or in other words, so that the diagram in Fig. 2 commutes for all k.

A commutative diagram is *trivial* if it remains satisfied if we replace f_k by any random permutation π. Each non-trivial commutative diagram for f is an interesting local property of the round function.

Such local properties can be pieced together to obtain global properties by exploiting the compositional behavior of commutative diagrams. Refer to Fig. 3. If both small squares commute (i.e., if $\rho' \circ f_{k_1} = g \circ \rho$ and $\rho'' \circ f_{k_2} = g' \circ \rho'$), then whole diagram commutes (e.g., $\rho'' \circ f_{k_2} \circ f_{k_1} = g' \circ g \circ \rho$). In other words, if ρ, ρ' form a local property of the first round f_{k_1} and if ρ', ρ'' form a local property of the second round f_{k_1}, then ρ, ρ'' form a global property of the first two rounds $f_{k_2} \circ f_{k_1}$.

Note that the requirement is that we have a local property of each round, *and* that the local properties match up appropriately. If ρ, ρ' is a local property of the first round and φ, φ' is a local property of the second round, these two

$$\begin{tikzcd}
\mathcal{M} \arrow[r, "\rho"] \arrow[d, "f_{k_1}"'] & Y \arrow[d, "g"] \\
\mathcal{M} \arrow[r, "\rho'"] \arrow[d, "f_{k_2}"'] & Y' \arrow[d, "g'"] \\
\mathcal{M} \arrow[r, "\rho''"] & Y''
\end{tikzcd}$$

Fig. 3. Splicing together a local property for f_{k_1} and a local property for f_{k_2} to obtain a property of their composition.

can only be composed if $\rho' = \varphi$. Thus there is a compatibility requirement that must be satisfied before two local properties can be composed.

The same kind of reasoning can be extended inductively to obtain a global property of the cipher as a whole. The requirement is that we obtain local properties for the rounds that are compatible. See Fig. 4. If each local property is non-trivial, then the global property so obtained will be non-trivial.

Any non-trivial global property for the cipher as a whole immediately leads to a distinguishing attack. Suppose we have ρ, ρ', g so that $\rho' \circ E_k = g \circ \rho$ holds for all k. Then our distinguishing attack is straightforward: we obtain a few known-plaintext/ciphertext pairs (x_i, y_i) and we check whether $\rho'(y_i) \overset{?}{=} g(\rho(x_i))$ holds for all of them. When the known texts are obtained from the real cipher (E_k), these equalities will always hold. However, since our property is non-trivial, the equalities will not always hold if the pairs (x_i, y_i) were obtained from an ideal cipher (π, a random permutation). The distinguishing advantage of such an attack depends on the details of the projections chosen, but we can typically expect to obtain a significant attack.

Example: Madryga. As a concrete example of a commutative diagram attack, let us examine Madryga, an early cipher design. Eli Biham discovered that the Madryga round function preserves the parity of its input. Hence, we may choose the parity function as our projection $\rho : \{0,1\}^{64} \to \{0,1\}$, i.e., $\rho(x_1, \ldots, x_{64}) = x_1 \oplus \cdots \oplus x_{64}$. When taken with the identity function, we obtain a global property for Madryga, as depicted in Fig. 5. This yields a distinguishing attack on Madryga with advantage $1/2$, as $\Pr[\rho(E_k(x)) = \rho(x)] = 1$ yet $\Pr[\rho(\pi(x)) = \rho(x)] = 1/2$.

Commutative diagrams are mathematically elegant. However, they are not, on their own, powerful enough to successfully attack many ciphers; another idea is needed. We shall describe next how these ideas may be extended to model statistical attacks, which turn out to be significantly more powerful.

$$\begin{tikzcd}
\mathcal{M} \arrow[r, "\rho_0"] \arrow[d, "f_{k_1}"'] & Y_0 \arrow[d, "g_1"] \\
\mathcal{M} \arrow[r, "\rho_1"] \arrow[d, dotted] & Y_1 \arrow[d, dotted] \\
\mathcal{M} \arrow[r, "\rho_{n-1}"] \arrow[d, "f_{k_n}"'] & Y_{n-1} \arrow[d, "g_n"] \\
\mathcal{M} \arrow[r, "\rho_n"] & Y_n
\end{tikzcd}$$

Fig. 4. Splicing together local properties for each round to obtain a global property for E_k, the cipher as a whole.

$$\begin{tikzcd}
\{0,1\}^{64} \arrow[r, "\rho"] \arrow[d, "E_k"'] & \{0,1\} \arrow[d, "\mathrm{id}"] \\
\{0,1\}^{64} \arrow[r, "\rho"] & \{0,1\}
\end{tikzcd}$$

Fig. 5. A global property of the Madryga block cipher. Here ρ is the parity function.

4 Statistical attacks

We now turn our attention to statistical attacks. The natural idea is to look at diagrams that only commute with some probability.

A reasonable first attempt might be to introduce the notion of *probabilistic commutative diagrams*. See Fig. 6, which is intended to show a diagram that commutes with probability p. In other words, though the relation $\rho' \circ f_k = g \circ \rho$ does not hold, we do have

$$\Pr_{X \leftarrow \mathcal{M}}[\rho'(f_k(X)) = g(\rho(X))] = p$$

for all $k \in \mathcal{K}$, where here the probability is taken with respect to the choice of X uniformly at random from $\mathcal{M}$. Written informally: $\rho' \circ f_k = g \circ \rho$ holds with probability p. (We could easily imagine many variants of this definition, for instance, by taking the probability over both the choice of X and k; however, we will not pursue such possibilities here.)

$$\begin{tikzcd}
\mathcal{M} \arrow[r, "\rho"] \arrow[d, "f_k"'] \arrow[dr, phantom, "\text{prob. } p"] & Y \arrow[d, "g"] \\
\mathcal{M} \arrow[r, "\rho'"] & Y'
\end{tikzcd}$$

Fig. 6. A local property that holds with probability p.

Probabilistic commutative diagrams share many useful properties with their deterministic cousins. First, probabilistic commutative diagrams can be composed. Suppose projections ρ, ρ' form a local property with prob. p for the first round, and ρ', ρ'' form a local property with prob. p' for the second round. Then ρ, ρ'' form a property for the composition of the first two rounds, and assuming our cipher is a Markov cipher [9], the composed property holds with prob. at least $p \cdot p'$. Second, probabilistic commutative diagrams for the whole cipher usually lead to distinguishing attacks. Suppose $\rho' \circ E_k = g \circ \rho$ holds with probability p, and $\rho' \circ \pi = g \circ \rho$ holds with probability q. Then there is a simple distinguishing attack that uses one known-plaintext/ciphertext pair (x, y) and has advantage $|p - q|$: we simply check whether $\rho'(y) \overset{?}{=} g(\rho(x))$.

Probabilistic commutative diagrams may appear fairly natural on first glance, but on further inspection, they seem to be lacking in some important respects. For our purposes, it will be useful to introduce a more general notion, which we term *stochastic commutative diagrams*. If we let the random variable X be uniformly distributed on $\mathcal{M}$, the maps ρ, ρ', E_k induce a Markov process on the pair of random variables $\rho(x)$, $\rho'(E_k(X))$. The associated transition matrix M is given by

$$M_{i,j} = \Pr_{X \leftarrow \mathcal{M}}[\rho'(E_k(X)) = j | \rho(X) = i].$$

Note that there is an implicit dependence on k, but for simplicity in this paper we will only consider the case where each key $k \in \mathcal{K}$ yields the same transition matrix M. (This corresponds to assuming that the Hypothesis of Stochastic Equivalence holds.) An example stochastic commutative diagram is shown pictorially in Fig. 7.

Stochastic commutative diagrams yield distinguishing attacks. Let M' be the transition matrix induced by the Markov process $\rho(X), \rho'(\pi(X))$, where π is a random permutation. Then our stochastic commutative diagram yields a distinguishing attack that uses one chosen plaintext query and achieves advantage $\frac{1}{2}||M - M'||_\infty$.

Linear cryptanalysis. Linear cryptanalysis may now be recognized as a special case of a stochastic commutative diagram attack. Suppose we have a ℓ-bit block cipher $E_k : \{0,1\}^\ell \to \{0,1\}^\ell$. In Matsui's linear cryptanalysis [10], the codebreaker somehow selects a pair of linear maps $\rho, \rho' : \{0,1\}^\ell \to \{0,1\}$, and

$$\begin{tikzcd}
\mathcal{M} \arrow[r, "\rho"] \arrow[d, "E_k"'] \arrow[dr, phantom, "\text{stochastic}"] & Y \arrow[d, "M"] \\
\mathcal{M} \arrow[r, "\rho'"] & Y'
\end{tikzcd}$$

Fig. 7. A stochastic commutative diagram for E_k. Here M is the transition matrix of the Markov process $\rho(X) - \rho'(E_k(X))$.

$$\begin{tikzcd}
\{0,1\}^\ell \arrow[r, "\rho"] \arrow[d, "E_k"'] \arrow[dr, phantom, "\text{stochastic}"] & \{0,1\} \arrow[d, "M"] \\
\{0,1\}^\ell \arrow[r, "\rho'"] & \{0,1\}
\end{tikzcd}$$

Fig. 8. A formulation of linear cryptanalysis as a stochastic commutative diagram attack. Here ρ and ρ' are linear maps.

then we use the stochastic commutative diagram shown in Fig. 8. For instance, the linear characteristic $\Gamma \to \Gamma'$ corresponds to the projections $\rho(x) = \Gamma \cdot x$, $\rho'(x) = \Gamma' \cdot x$.

In a linear attack, we obtain a 2×2 transition matrix M of the form

$$M = \begin{bmatrix} \frac{1}{2} + \frac{\epsilon}{2} & \frac{1}{2} - \frac{\epsilon}{2} \\ \frac{1}{2} - \frac{\epsilon}{2} & \frac{1}{2} + \frac{\epsilon}{2} \end{bmatrix}. \tag{1}$$

The transition matrix associated to a random permutation π is U, the 2×2 matrix where all entries are $1/2$. Therefore, the distinguishing advantage of a linear cryptanalysis attack using one known text is $\frac{1}{2}||M - U||_\infty = \epsilon/2$. It is not hard to verify that $\Theta(1/\epsilon^2)$ known texts suffice to obtain an attack with distinguishing advantage $1/2$ (say). Compare to Matsui's rule of thumb, which says that $8/(\epsilon/2)^2 = 32/\epsilon^2$ texts suffice.

Matsui's piling-up lemma can also be re-derived within this framework. Let M_1, M_2 be transition matrices for the first and second round, respectively, taking the form shown in Equation (1) albeit with ϵ replaced by ϵ_1, ϵ_2. It is not hard to verify that $M_1 M_2$ also takes the form shown in Equation (1), but with ϵ replaced by $\epsilon_1 \epsilon_2$. Hence $||M_1 M_2 - U||_\infty = ||M_1 - U||_\infty \times ||M_2 - U||_\infty$. This is exactly the piling-up lemma for computing the bias of a multi-round linear characteristic given the bias of the characteristic for each round.

After seeing this formulation of linear cryptanalysis, our use of the name "projection" to describe the maps ρ, ρ' can be justified as follows. Consider the vector subspace $\mathcal{V} = \{0, \Gamma\}$ of $\{0,1\}^\ell$. We obtain a canonical isomorphism of vector spaces $\mathcal{V} \cong \{0,1\}$. Then we can view $\rho : \{0,1\}^\ell \to \mathcal{V}$ as taking the form of a projection onto the subspace $\mathcal{V}$. In other words, we write $\{0,1\}^\ell$ as the direct sum $\{0,1\}^\ell = \mathcal{V} \oplus \mathcal{V}^T$, write each $x \in \{0,1\}^\ell$ as a sum $x = y \oplus z$ for $y \in \mathcal{V}$ and $z \in \mathcal{V}^T$, and then let $\rho(x) = y$ be the projection of x onto $\mathcal{V}$.

Mod n cryptanalysis. Notice that mod n attacks also fall within this framework. If $\mathcal{M} = \mathbb{Z}/2^\ell\mathbb{Z}$, we can use the projection $\rho : \mathbb{Z}/2^\ell\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ given by $\rho(x) = x \bmod n$. In this way we recover the mod n attack.

In general, if $\mathcal{M}$ is any abelian group with subgroup $S \subseteq \mathcal{M}$, we may consider projections of the form $\rho : \mathcal{M} \to \mathcal{M}/S$ given by $\rho(x) = x \bmod S$. Linear cryptanalysis is simply the special case where $\mathcal{M} = (\{0,1\}^\ell, \oplus)$ and S has index 2, and mod n cryptanalysis is the special case where $\mathcal{M} = (\mathbb{Z}, +)$ and $S = n\mathbb{Z}$.

Linear cryptanalysis with multiple approximations. Linear cryptanalysis with multiple approximations also falls naturally within this framework. Suppose we have a list of masks $\Gamma_1, \Gamma_2, \ldots, \Gamma_d \in \{0,1\}^\ell$, and assume that these masks are linearly independent as vectors in $\{0,1\}^\ell$. Let $\mathcal{V}$ be the vector subspace of dimension d spanned by $\Gamma_1, \ldots, \Gamma_d$, i.e., $\mathcal{V} = \{0, \Gamma_1, \Gamma_2, \Gamma_1 \oplus \Gamma_2, \ldots\}$. Choose the canonical isomorphism $\mathcal{V} \cong \{0,1\}^d$, i.e., $\sum_i c_i \Gamma_i \mapsto (c_1, \ldots, c_d)$. We can define the projection $\rho : \{0,1\}^\ell \to \{0,1\}^d$ by $\rho(x) = (\Gamma_1 \cdot x, \ldots, \Gamma_d \cdot x)$, or equivalently, as the projection $\rho : \{0,1\}^\ell \to \mathcal{V}$ from $\{0,1\}^\ell$ onto the subspace $\mathcal{V}$. We can build ρ' from $\Gamma_1', \ldots, \Gamma_d'$ similarly. Then, we consider the Markov process M induced by E_k, ρ, ρ'. The distinguishing advantage of this attack is given by $\frac{1}{2}||M - M'||_\infty$, as before, except that now we are working with $2^d \times 2^d$ matrices. Notice that these d-bit projections ρ, ρ' simultaneously capture all 2^{2d} linear approximations of the form $\Gamma \to \Gamma'$ for some $\Gamma \in \mathcal{V}$, $\Gamma' \in \mathcal{V}'$, so this is a fairly powerful attack.

5 Higher-order attacks

The next idea is to examine plaintexts two (or more) at a time. If $f : \mathcal{M} \to \mathcal{M}$ is any function, let $\hat{f} : \mathcal{M}^2 \to \mathcal{M}^2$ be defined by $\hat{f}(x, x') = (f(x), f(x'))$. More generally, we can take $\hat{f} : \mathcal{M}^d \to \mathcal{M}^d$ and $\hat{f}(x_1, \ldots, x_d) = (f(x_1), \ldots, f(x_d))$ for any fixed d; this is known as a d-th order attack. Then, to distinguish E_k from π, the idea is to use stochastic commutative diagrams that separate $\hat{E}_k$ from $\hat{\pi}$.

Complementation properties. Complementation attacks form a simple instance of a higher-order attack. Suppose there is some Δ so that $E_k(x \oplus \Delta) = E_k(x) \oplus \Delta$. Then we can define $\rho : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ by $\rho(x, x') = x - x'$, and we obtain the diagram shown in Fig. 9. Note that in this case the existence of the complementation property implies that $M_{\Delta,\Delta} = 1$. If M' denotes the transition matrix induced by an ideal cipher (namely, π), then $M'_{\Delta,j} = 1/(|\mathcal{M}| - 1)$ for each $j \neq 0$, and so $||M - M'||_\infty \geq 1 - 1/(|\mathcal{M}| - 1) + (|\mathcal{M}| - 2)/(|\mathcal{M}| - 1) = 2 - 2/(|\mathcal{M}| - 1)$. In other words, there is an attack using only 2 chosen plaintexts and achieving distinguishing advantage $1 - 1/(|\mathcal{M}| - 1)$.

$$\begin{tikzcd}
\mathcal{M}\times\mathcal{M} \arrow[r, "\rho"] \arrow[d, "\hat{E}_k"'] \arrow[dr, phantom, "\text{stochastic}"] & \mathcal{M} \arrow[d, "M"] \\
\mathcal{M}\times\mathcal{M} \arrow[r, "\rho"] & \mathcal{M}
\end{tikzcd}$$

Fig. 9. The basis of differential-style attacks, as a commutative diagram. We use the projection $\rho(x, x') = x - x'$.

Differential cryptanalysis. A natural extension is to generalize the above attack by looking for some matrix element $M_{\Delta,\Delta'}$ with surprisingly large probability, rather than looking for a matrix element with probability 1. Indeed, such a modification yields exactly Biham & Shamir's differential cryptanalysis [2], and any large matrix element $M_{\Delta,\Delta'}$ gives us a differential $\Delta \to \Delta'$ with probability $p = M_{\Delta,\Delta'}$. Notice that when $p \gg 1/(|\mathcal{M}| - 1)$, we have $\frac{1}{2}||M - M'||_\infty \geq p - 1/(|\mathcal{M}| - 1)$, hence with 2 chosen plaintexts we obtain an attack with advantage $\approx p$. One can readily verify that with $2m$ chosen plaintexts, the iterated attack has advantage $1-(1-1/p)^m-(1-1/(|\mathcal{M}|-1))^m$, which is $\approx 1-1/e$ for $m = 1/p$. Thus, $2/p$ chosen plaintexts suffice to distinguish with good advantage.

Also, our framework easily models differential cryptanalysis with respect to other groups. For instance, when using additive differentials in the group $\mathcal{M} = (\mathbb{Z}/2^{64}\mathbb{Z}, \boxplus)$, we can choose the projection $\rho : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ given by $\rho(x, x') = x - x' \bmod 2^{64}$.

Impossible differential cryptanalysis. Alternatively, we could look for matrix elements in M that are surprisingly small. If we look for entries that have probability 0, say $M_{\Delta,\Delta'} = 0$, then we obtain the impossible differential attack. In this case the differential $\Delta \to \Delta'$ will be impossible (it can never happen for the real cipher E_k). This yields an attack that can distinguish with good advantage once about $|\mathcal{M}|$ texts are available to the attacker.

Differential-linear cryptanalysis. Write $E_k = f' \circ f$. In a differential-linear attack [5], one covers the first half of the cipher (f) with a differential characteristic and approximates the second half of the cipher (f') with a linear characteristic. This can be modeled within our framework as follows. Our development so far suggests we should use the projections $\rho, \rho' : \mathcal{M}^2 \to \mathcal{M}$, $\rho(x, x') = \rho'(x, x') = x - x'$ to cover f and projections $\eta, \eta' : \mathcal{M} \to \{0, 1\}$ to cover f'. However, in this case, we will not be able to match η up with ρ', because neither their domains nor their ranges agree.

The solution is to introduce functions $\hat{\eta}, \hat{\eta}' : \mathcal{M}^2 \to \{0, 1\}$ given by $\hat{\eta}(x, x') = \eta(x - x')$ and $\hat{\eta}'(x, x') = \eta'(x - x')$. Suppose our differential characteristic has probability $p \gg 1/|\mathcal{M}|$, or in other words, ρ, ρ' commute with f with probability p. Then $\rho, \hat{\eta}$ will usually commute with $\hat{f}$ with non-trivial probability (heuristically, about $\frac{1}{2} + \frac{p}{2}$, though this is not guaranteed). Likewise, suppose our linear characteristic holds with probability $\frac{1}{2} \pm \frac{\epsilon}{2}$, or in other words, η, η' commute with f' with probability $\frac{1}{2} \pm \frac{\epsilon}{2}$. Then $\hat{\eta}, \hat{\eta}'$ will form a linear approximation for $\hat{f}'$ with probability $\frac{1}{2} \pm \frac{\epsilon^2}{2}$. These two properties can be composed to obtain a property $\rho, \hat{\eta}'$ for the whole cipher, typically with probability $\frac{1}{2} \pm \frac{p\epsilon^2}{2}$.

Hence a differential-linear attack with two chosen texts will typically have distinguishing advantage roughly $\frac{1}{2}p\epsilon^2$. Consequently, our framework predicts that such a cipher can be broken with $\Theta(1/p\epsilon^2)$ chosen texts. This corresponds closely to the classical estimate [1].

Higher-order differential cryptanalysis. Higher-order differentials [7] can also be modeled within our framework. Let us give a simple example. If $f(X)$ is a polynomial of degree 2, then $f(X+\Delta_0+\Delta_1) - f(X+\Delta_0) + f(X+\Delta_1) - f(X)$ is a constant polynomial, giving a way to distinguish f from random with 4 chosen plaintexts. This corresponds to choosing 4th order projections $\rho(w, x, y, z) = (x - w, y - w, z - x - y + w)$ and $\rho'(w, x, y, z) = z - x + y - w$, deriving a transition matrix M, and noticing that we have a matrix entry $M_{(\Delta_0,\Delta_1,0),\Delta'}$ whose value is 1 for the real cipher but much smaller for a random permutation. More generally, if f is a polynomial of degree d, then the dth order differential of f is a constant, and the $d+1$-th order differential is zero. Such higher order differential attacks can likewise be expressed be expressed as a higher-order commutative diagram attack.

Truncated differential cryptanalysis. Truncated differential attacks [8] also fit within our framework. Given a block cipher $E_k : \{0,1\}^\ell \to \{0,1\}^\ell$, we choose projections of the form $\rho : \{0,1\}^\ell \times \{0,1\}^\ell \to \{0,1\}^m$, where $\rho(x, x') = \varphi(x - x')$ and $\varphi : \{0,1\}^\ell \to \{0,1\}^m$ is an appropriately chosen linear map.

Then we can look for an entry $M_{\Delta,\Delta'}$ in the transition matrix so obtained that has surprisingly large probability, and this will correspond to a truncated differential $\Delta \to \Delta'$ of the same probability. This truncated differential corresponds to the class of $2^{2\ell-2m}$ conventional differentials $\delta \to \delta'$, where $\delta \in \varphi^{-1}(\Delta)$ and $\delta' \in \varphi^{-1}(\Delta')$. Alternately, we can look for an entry with surprisingly low probability, and this yields an impossible (or improbable) truncated differential that can be used in an attack.

In most truncated differential attacks, the linear map φ simply ignores part of the block. For instance, the truncated difference $(a, 0, 0, b)$ (where a, b are arbitrary differences) might correspond to the linear map $\varphi(w, x, y, z) = (x, y)$ and the projected value $\Delta = (0, 0)$. However, for maximum generality, we allow φ to be chosen as any linear map whatsoever.

The above account of truncated differentials is slightly naive. It leads to very large matrices, because a truncated difference of the form (say) $(a, 1, 2, b)$ is distinguished from the truncated difference $(a, 1, 3, b)$. However, in practice it is more common for cryptanalysts to care only about distinguishing between zero and non-zero words, with little reason to make any distinction between the different non-zero values.

Fortunately, our treatment can be amended to better incorporate typical cryptanalytic practice, as follows. Consider, as a concrete example, truncated differential attacks on Skipjack, where the block is $\mathcal{M} = \{0,1\}^{64}$ and where attacks typically look at which of the four 16-bit words of the difference are zero or not. Consider the following 67 vector subspaces of $\{0,1\}^{64}$: $\{0\}$, $\{(a,0,0,0) : a \in \{0,1\}^{16}\}$, $\{(0,a,0,0) : a \in \{0,1\}^{16}\}$, …, $\{(a,a,0,0) : a \in \{0,1\}^{16}\}$, $\{(a,0,a,0) : a \in \{0,1\}^{16}\}$, …, $\{(a,b,0,0) : a,b \in \{0,1\}^{16}\}$, $\{(a,0,b,0) : a,b \in \{0,1\}^{16}\}$, …, $\{(a,b,c,d) : a,b,c,d \in \{0,1\}^{16}\}$. These can be put into one-to-one correspondence with the 67 vector subspaces of $\{0,1\}^4$ in a natural way. Moreover, to any block $x \in \{0,1\}^{64}$ we can associate its characteristic vector $(\chi_1, \ldots, \chi_{67})$, where χ_i is 1 if x is in the i-th subspace and 0 otherwise. This induces an equivalence relation $\sim$ on $\{0,1\}^{64}$, where two blocks are considered equivalent if they have the same characteristic vector. We can now consider the projection $\rho : \{0,1\}^{64} \to \{0,1\}^{64}/\sim$ that maps x to its equivalence class under $\sim$. In this way we obtain a 67×67 transition matrix M that captures the probability of all 67^2 word-wise truncated differentials for Skipjack [11]. A similar construction can be used for ciphers of other word and block lengths. This leads to smaller transition matrices and a more satisfying theory of truncated differential cryptanalysis.

Generalized truncated differential cryptanalysis. Armed with these ideas, we can now propose a new attack not previously seen in the literature. We retain the basic set-up from truncated differential attacks, but we replace the probabilistic commutative diagrams with stochastic commutative diagrams. In other words, instead of looking for a single entry in the transition matrix with unusually large (or small) probability, we use the matrix norm $\frac{1}{2}||M - M'||_\infty$. This approach allows us to exploit many small biases spread throughout the matrix M, rather than being confined to only taking advantage of one bias and ignoring the rest.

This may look like a very small tweak; however, it contributes considerable power to the attack. As we shall see, it subsumes all the previous attacks as special cases of generalized truncated differentials. The ability to unify many existing attacks, and to generate new attacks, using such a simple and natural-looking extension to prior work is one of the most striking features of our framework.

Generalized truncated differentials generalize conventional differential and impossible differential attacks. If there is a single entry in M with unusually large (or small) probability, then the matrix norm $\frac{1}{2}||M - M'||_\infty$ will also be large. Hence, the existence of a conventional differential or impossible differential attack with advantage ϵ implies the existence of a generalized truncated attack with the same advantage ϵ.

Likewise, linear cryptanalysis can also be viewed as a special case of generalized truncated differentials. As we argued before, if $\eta, \eta' : \{0,1\}^\ell \to \{0,1\}$ form a linear characteristic with probability $\frac{1}{2} \pm \frac{\epsilon}{2}$, then $\hat{\eta}, \hat{\eta}'$ given by $\hat{\eta}(x,x') = \eta(x - x')$, etc., has probability $\frac{1}{2} \pm \frac{\epsilon^2}{2}$. It may appear that we have diluted the power of the attack, because the distinguishing advantage has decreased from $\epsilon/2$ (for one known text in a linear attack) to $\epsilon^2/2$ (for one pair of texts in a generalized truncated attack). However, this is offset by an increase in the number of pairs of texts available in a generalized truncated attack: given a pool of n known texts, one can form n^2 pairs of texts. These two factors turn out to counterbalance each other. If k out of n texts follow η, η', then $k^2 + (n-k)^2$ out of n^2 pairs follow $\hat{\eta}, \hat{\eta}'$. Note that $h(k) = k^2 + (n-k)^2$ is a strictly increasing function of k, for $k \geq n/2$. Hence for any threshold T in a linear attack, there is a corresponding threshold $h(T)$ that makes the generalized truncated differential attack work with the same number of known texts and roughly the same distinguishing advantage. Consequently, the existence of a linear attack implies the existence of a generalized truncated attack with about the same advantage.

Differential-linear attacks are also subsumed by generalized truncated differentials. Because both a differential and a linear characteristic can be viewed as a generalized truncated differential, they can be concatenated. In a differential-linear attack, the transition is abrupt and binary, but generalized truncated attacks allow to consider other attacks, for instance with a gradual transition between differential- and linear-style analysis, or with hybrids partway between differential and linear attacks.

Intuitively, the power of generalized truncated differential attacks comes from the extra degrees of freedom available to the cryptanalyst. In a differential attack, the cryptanalyst can freely choose which matrix entry $M_{\Delta,\Delta'}$ to focus on, but has little control over ρ, ρ'. In a linear attack, the cryptanalyst can freely choose ρ, ρ' in some clever way, but has no choice over the matrix M. A generalized truncated differential attack allows the cryptanalyst to control both aspects of the attack at the same time.

6 Algebraic attacks

One noticeable trend over the past few decades is that more and more block cipher designs have come to incorporate algebraic structure. For instance, the AES S-box is based on inversion in the field $GF(2^8)$. Yet this brings an opportunity for attacks that exploit this structure, and a rich variety of algebraic cryptanalytic methods have been devised: interpolation attacks, rational interpolation, probabilistic interpolation, and so on.

Interpolation attacks. The basic interpolation attack [7] is easy to understand. We express each round f_{k_i} as a polynomial $q_{k_i}(X) \in \mathbb{F}[X]$ over some field $\mathbb{F}$, and in this way we obtain the commutative diagram shown in Fig. 10. Notice that composition of commutative diagrams allows to express the whole cipher as a polynomial $q_k(X) = q_{k_n}(\cdots(q_{k_2}(q_{k_1}(X)))\cdots)$. The polynomial $q_k(X)$ may depend on the key k in some possibly complex way, hence the attacker usually will not know $q_k(X)$ a priori. Consequently, a distinguishing attack based on this property must work a little differently.

The standard interpolation attack exploits the fact that $d+1$ points suffice to uniquely determine a polynomial of degree d. Given $\deg q_k(X) + 1$ known plaintext/ciphertext pairs for E_k, we can reconstruct the polynomial $q_k(X)$ using Lagrange interpolation (for instance), and then check one or two additional

$$\begin{tikzcd}
\mathbb{F} \arrow[r, "\mathrm{id}"] \arrow[d, "f_{k_1}"'] & \mathbb{F} \arrow[d, "q_{k_1}"] \\
\mathbb{F} \arrow[r, "\mathrm{id}"] \arrow[d, dotted] & \mathbb{F} \arrow[d, dotted] \\
\mathbb{F} \arrow[r, "\mathrm{id}"] \arrow[d, "E_{k_n}"'] & \mathbb{F} \arrow[d, "q_{k_n}"] \\
\mathbb{F} \arrow[r, "\mathrm{id}"] & \mathbb{F}
\end{tikzcd}$$

Fig. 10. The basis of an interpolation attack. Here each $q_{k_i}(X)$ is a polynomial over the field $\mathbb{F}$.

known texts for consistency with the recovered polynomial. This allows to distinguish any cipher with this property from a random permutation.

The idea can be generalized in many ways. We need not restrict ourself to univariate polynomials; we can generalize to multivariate polynomials $\mathbf{q}(\mathbf{X})$, where $\mathbf{X} = (X_1, \ldots, X_m)$ represents a vector of m unknowns, and where $\mathbf{q}(\mathbf{X}) = (q_1(\mathbf{X}), \ldots, q_m(\mathbf{X}))$ represents a vector of m multivariate polynomials. In this case, the number of texts needed corresponds to the number of coefficients of q_k not known to be zero.

Also, one can naturally derive statistical versions of interpolation cryptanalysis by replacing the commutative diagram in Fig. 10 with a probabilistic commutative diagram with some probability p. Then noisy polynomial reconstruction techniques (e.g., list decoding of Reed-Solomon codes) will allow us to mount a distinguishing attack.

We can also use meet-in-the-middle techniques, using the polynomial q_k to cover the first half of the cipher and q'_k to cover the remaining rounds. Notice how these generalizations come naturally under the commutative diagram framework.

Rational interpolation attacks. Another generalization is the notion of rational interpolation attacks. If $q(X), q'(X)$ are any two polynomials with no common factor and where $q'(X)$ is not the zero polynomial, then $r(X) = q(X)/q'(X)$ is called a *rational polynomial.* It is then natural to consider the variant on the commutative diagram in Fig. 10 where each polynomial q_{k_i} is replaced by a rational polynomial r_{k_i}.

Note that rational polynomials are closed under composition, hence we can derive a global approximation for the whole cipher from rational approxima-

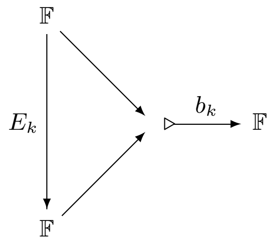

Fig. 11. The basis of a bivariate interpolation attack. Here $b_k(X, Y)$ represents a bivariate polynomial over $\mathbb{F}$, and the intended interpretation of the diagram is that $b_k(x, E_k(x)) = 0$ for all $x \in \mathbb{F}$.

tions of the individual round functions. If we can express the cipher as a rational polynomial $E_k(x) = q_k(x)/q'_k(x)$, and if we have a supply of known plaintext/ciphertext pairs (x_i, y_i), then we obtain the equations $y_i \cdot q'_k(x_i) = q_k(x_i)$. Linear algebra reveals the rational polynomial $q_k(X)/q'_k(X)$, which gives us a distinguishing attack on the cipher.

Bivariate interpolation. The notion of interpolation and rational interpolation can be generalized to obtain what might be called a *bivariate interpolation attack.* The idea is to seek a family of bivariate polynomials $b_k(X, Y) \in \mathbb{F}[X, Y]$ so that $b_k(x, f_k(x)) = 0$ for all $x \in \mathcal{M}$ and all $k \in \mathcal{K}$. This gives a local property of the round function f.

Local bivariate properties can be composed to obtain a global property for the whole cipher. Suppose the first round satisfies a bivariate relation $b(x, f(x)) = 0$ and the second round satisfies $b'(y, f'(y)) = 0$. Define

$$b''(X, Z) = \mathrm{Res}_Y(b(X, Y), b'(Y, Z)).$$

Then the composition of the two rounds will satisfy the relation $b''(x, f'(f(x))) = 0$ for all $x \in \mathbb{F}$. The latter follows from a property of the resultant: given $f(Y), g(Y) \in \mathcal{R}[Y]$, the resultant $\mathrm{Res}_Y(f(Y), g(Y))$ is a value in $\mathcal{R}$, and if f, g share a common root over $\mathcal{R}$, then the resultant will be zero. Letting $\mathcal{R} = \mathbb{F}[X, Z]$, $f(Y) = b(X, Y)$, and $g(Y) = b'(Y, Z)$ verifies the claimed result about $b''(X, Z)$.

In this way, we can compose bivariate relations for each round to obtain a bivariate relation for the whole cipher. Once we have a bivariate relation $b_k(x, E_k(x)) = 0$ for the whole cipher, we can use polynomial interpolation to reconstruct b_k given a sufficient quantity of known plaintext/ciphertext pairs. Unfortunately, in general the degree of the bivariate polynomial for the whole cipher can grow rapidly as the number of rounds increases.

Notice that interpolation attacks fall out as a special case of bivariate interpolation. If the first and second rounds of the cipher can be expressed as polynomials $q(X), q'(Y)$, this induces bivariate relations $b(X, Y) = q(X) - Y$

and $b'(Y,Z) = q'(Y) - Z$. Taking the resultant yields

$$b''(X,Z) = \mathrm{Res}_Y(b(X,Y), b(Y,Z)) = \mathrm{Res}_Y(q(X)-Y, q'(Y)-Z) = q'(q(X))-Z,$$

which is nothing more than a round-about derivation of the obvious fact that the composition of first and second rounds may be expressed by the polynomial $q'(q(X))$.

Likewise, rational interpolation attacks are a special case of bivariate interpolation. Suppose the first and second rounds can be expressed as rational polynomials $p(X)/p'(X)$ and $q(Y)/q'(Y)$. We obtain the bivariate relations $b(X,Y) = p'(X) \cdot Y - p(X)$ and $b'(Y,Z) = q'(Y) \cdot Z - q(Y)$. Taking the resultant yields a bivariate relation for the composition of the first two rounds.

Probabilistic bivariate attacks have actually been suggested before by Jakobsen [6] and applied by others to DES [12], but it was not previously explained how to compose local approximations to obtain global approximations, nor was it noticed that bivariate attacks generalize and unify interpolation and rational interpolation.

7 Discussion

Closure properties. The common theme here seems to be that closure properties enable cryptanalysis. For instance, differential and linear attacks exploit the fact that the set of linear functions is closed under composition: $g \circ f$ is linear if f, g are. Likewise for the set of polynomials, of rational polynomials, and so on. More generally, we may form a norm $\mathcal{N}(\cdot)$ on functions that grows slowly under composition and that corresponds somehow to the cost of an attack. Consider interpolation attacks: letting $\mathcal{N}(q) = \deg q$ for polynomials $q(X)$, we find $\mathcal{N}(g \circ f) \leq \mathcal{N}(f) \times \mathcal{N}(g)$, hence if we can find low-degree properties for individual round functions, the corresponding global property for the whole cipher will have not-too-large degree. Perhaps other ways to place a metric space structure on the set of bijective functions $f : \mathcal{M} \to \mathcal{M}$ will lead to other cryptanalytic advances in the future.

Related work. Commutative diagram cryptanalysis draws heavily on ideas found in previous frameworks, most notably Vaudenay's chi-squared cryptanalysis [13] and Harpes' partitioning cryptanalysis [4]. Vaudenay's work used linear projections $\rho, \rho' : \{0,1\}^\ell \to \{0,1\}^m$, and then applied the χ^2 statistical test to the pair $(\rho(X), \rho'(E_k(X))$. It turns out that the power of the χ^2 test is closely related to the ℓ_2 norm, $||M - M'||_2$, hence chi-squared cryptanalysis can be viewed as a variant of stochastic commutative diagrams where a different matrix norm is used. Partitioning cryptanalysis generalized this to allow arbitrary (not necessarily linear) projections ρ, ρ'.

We borrowed methods from Vaudenay's decorrelation theory [14] to calculate the distinguishing advantage of our statistical attacks. Also, Vaudenay shows how to build ciphers with provable resistance against all non-adaptive d-limited attacks, which corresponds to security against dth order commutative diagram attacks.

This work builds on an enormous quantity of work in the block cipher literature; it is a synthesis of many ideas that have previously appeared elsewhere. Due to space limitations, we have been forced to omit mention of a great deal of relevant prior work, and we apologize for all omissions.

8 Conclusion

We have introduced commutative diagram cryptanalysis and shown how it provides a new perspective on many prior attacks in the block cipher literature. We also described two new attack methods, generalized truncated differential cryptanalysis and bivariate interpolation, and demonstrated how they generalize and unify many previous attacks. It is an interesting open problem to extend this framework to incorporate more attacks, to discover more new attacks, or to build fast ciphers that are provably secure against commutative diagram cryptanalysis.

References

1. E. Biham, O. Dunkelman, N. Keller, “Enhancing Differential-Linear Cryptanalysis,” *ASIACRYPT 2002*, Springer-Verlag, LNCS 2501, pp.254–266.
2. E. Biham, A. Shamir, *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.
3. D. Coppersmith, S. Halevi, C. Jutla, “Cryptanalysis of Stream Ciphers with Linear Masking,” *CRYPTO 2002*, Springer-Verlag, LNCS 2442, pp.515–532.
4. C. Harpes, J.L. Massey, “Partitioning Cryptanalysis,” *FSE ’97*, Springer-Verlag, LNCS 1267, pp.13–27.
5. M.E. Hellman, S.K. Langford, “Differential-linear cryptanalysis,” *CRYPTO ’94*, Springer-Verlag, LNCS 839, pp.26–39.
6. T. Jakobsen, “Cryptanalysis of Block Ciphers with Probabilistic Non-Linear Relations of Low Degree,” *CRYPTO ’98*, Springer-Verlag, LNCS 1462, pp.212–222.
7. T. Jakobsen, L.R. Knudsen, “Attacks on Block Ciphers of Low Algebraic Degree,” *J. Cryptology*, 14(3):197–210, 2001.
8. L. Knudsen, “Truncated and Higher Order Differentials,” *FSE ’94*, Springer-Verlag, LNCS 1008, pp.196–211.
9. X. Lai, J. Massey, S. Murphy, “Markov Ciphers and Differential Cryptanalysis,” *EUROCRYPT ’91*, Springer-Verlag, LNCS 547, pp.17–38.
10. M. Matsui, “Linear Cryptanalysis Method for DES Cipher,” *EUROCRYPT ’93*, Springer-Verlag, LNCS 765, pp.386–397.
11. B. Reichardt, D. Wagner, “Markov truncated differential cryptanalysis of Skipjack,” *SAC 2002*, Springer-Verlag, LNCS 2595, pp.110–128.
12. T. Shimoyama, T. Kaneko, “Quadratic Relation of S-box and Its Application to the Linear Attack of Full Round DES,” *CRYPTO ’98*, Springer-Verlag, LNCS 1462, pp.200–211.
13. S. Vaudenay, “An Experiment on DES: Statistical Cryptanalysis,” *ACM CCS ’96*, ACM Press, pp.139–147.
14. S. Vaudenay, “Decorrelation: A Theory for Block Cipher Security,” *J. Cryptology*, Springer-Verlag, 16(4):249–286, Sept. 2003.