A Methodology for Differential-Linear Cryptanalysis and Its Applications[⋆]
(Extended Abstract)

Jiqiang Lu

Institute for Infocomm Research,
Agency for Science, Technology and Research
1 Fusionopolis Way, #19-01 Connexis, Singapore 138632
lvjiqiang@hotmail.com,jlu@i2r.a-star.edu.sg

Abstract. In 1994 Langford and Hellman introduced a combination of differential and linear cryptanalysis under two default independence assumptions, known as differential-linear cryptanalysis, which is based on the use of a differential-linear distinguisher constructed by concatenating a linear approximation with a (truncated) differential with probability 1. In 2002, by using an additional assumption, Biham, Dunkelman and Keller gave an enhanced version that can be applicable to the case when a differential with a probability of smaller than 1 is used to construct a differential-linear distinguisher. In this paper, we present a new methodology for differential-linear cryptanalysis under the original two assumptions implicitly used by Langford and Hellman, without using the additional assumption of Biham et al. The new methodology is more reasonable and more general than Biham et al.'s methodology, and apart from this advantage it can lead to some better differential-linear cryptanalytic results than Biham et al.'s and Langford and Hellman's methodologies. As examples, we apply it to attack 10 rounds of the CTC2 block cipher with a 255-bit block size and key, 13 rounds of the DES block cipher, and 12 rounds of the Serpent block cipher. The new methodology can be used to cryptanalyse other block ciphers, and block cipher designers should pay attention to this new methodology when designing a block cipher.

Key words: Block cipher, CTC2, DES, Serpent, Differential cryptanalysis, Linear cryptanalysis, Differential-linear cryptanalysis.

1 Introduction

Differential cryptanalysis was introduced in 1990 by Biham and Shamir [8]. Linear cryptanalysis was introduced in 1992 by Matsui and Yamagishi [31]. A differ-

[⋆] An earlier version of this work appeared in 2010 as part of Cryptology ePrint Archive Report 2010/025 [28], which was done when the author was with Eindhoven University of Technology (The Netherlands) under the support of the Dutch Sentinels project PINPASJC (No. TIF.6687).

Table 1. Our and previous main cryptanalytic results on CTC2 and DES

Cipher	Attack Technique	Rounds	Data	Time	Success Rate	Source
CTC2 (255-bit version)	Algebraic [12]	6	4CP	2^{253}Enc.	not specified	[11]
	Differential	$7^{\dagger}$	2^{15}CP	2^{15}Enc.	not specified	[15]
	Differential-linear	$8^{\dagger}$	2^{37}CP	2^{37}Enc.	61.8%	[15]
		10	2^{142}CP	2^{207}Enc.	99.9%	Section 5.4
DES	Differential	full	$2^{47.2}$CP	2^{37}Enc.	not specified	[10]
	Linear	full	2^{43}KP	2^{47}Enc.	85%	[29]
	Davis's attack [13]	full	2^{50}KP	2^{50}Enc.	51%	[4]
	Differential-linear	8	768CP	2^{40}Enc.	95%	[26]
		9	$2^{15.75}$CP	2^{38}Enc.	88.8%	[14]
		10	$2^{29.66}$CP	2^{44}Enc.	97%	Section 4.2
		13	$2^{52.1}$CP	$2^{54.2}$Enc.	99%	Section 4.2

†: There is a flaw; see Section 5.2 for detail.

ential cryptanalysis attack is based on the use of one or more so-called differentials, and a linear cryptanalysis attack is based on the use of one or more so-called linear approximations. Both the cryptanalytic methods were used to attack the full Data Encryption Standard (DES) [32] algorithm faster than exhaustive key search [10, 29].

In 1994 Langford and Hellman [26] introduced a combination of differential and linear cryptanalysis under two default independence assumptions, known as differential-linear cryptanalysis, and they applied it to break 8-round DES. Such an attack is constructed on a so-called differential-linear distinguisher; a differential-linear distinguisher treats a block cipher as a cascade of two sub-ciphers, and it uses a linear approximation for a sub-cipher and, for the other sub-cipher it uses a differential (or a truncated differential [22]) with a one probability that does not affect the bit(s) concerned by the input mask of the linear approximation. In 2002, by using an additional assumption Biham, Dunkelman and Keller [5] introduced an enhanced version of differential-linear cryptanalysis, which is applicable to the case when a differential with a smaller probability is used to construct a differential-linear distinguisher; and they applied the enhanced version to break 9-round DES. Differential-linear cryptanalysis has been used to yield the best currently published cryptanalytic results for a number of state-of-the-art block ciphers [5, 6, 15, 16].

In this paper, we present a new methodology for differential-linear cryptanalysis under the two default assumptions implicitly used by Langford and Hellman, without using the additional assumption due to Biham et al. The new methodology is more reasonable and more general than Biham et al.'s methodology, and it can lead to some better differential-linear cryptanalytic results than Biham et al.'s and Langford and Hellman's methodologies. As examples, we apply the new methodology to mount differential-linear attacks on 10 rounds of the CTC2 [11] block cipher with a 255-bit block size and key, 13 rounds of DES, and 12 rounds of the Serpent [1, 2] block cipher. In terms of the numbers of attacked rounds: The 10-round CTC2 attack is the first published cryptanalytic attack on the version of CTC2; the 13-round DES attack is much better than any previously published differential-linear cryptanalytic results for DES, though it is inferior to the best previously published cryptanalytic results for DES; and the 12-round Serpent attack matches the best previously published cryptanalytic result for Serpent, that was obtained under Biham et al.'s methodology. Due to page constraints, we will only present the attacks on CTC2 and DES in this paper, and give the attack on Serpent in the full version of this paper (which contains more material). Table 1 summarises both our and previous main cryptanalytic results on CTC2 and DES, where CP and KP refer respectively to the required numbers of chosen plaintexts and known plaintexts, and Enc. refers to the required number of encryption operations of the relevant version of CTC2 or DES.

The remainder of the paper is organised as follows. In the next section we give the notation used throughout the paper and briefly describe differential and linear cryptanalysis. In Section 3 we review Langford and Hellman's and Biham et al.'s methodologies and give our methodology for differential-linear cryptanalysis. In Sections 4–5 we present our cryptanalytic results on DES and CTC2, respectively. We discuss a few possible extensions to our methodology in Section 6. Section 7 concludes this paper.

2 Preliminaries

In this section we describe the notation, differential and linear cryptanalysis.

2.1 Notation

In the following descriptions, we assume that a number without a prefix is in decimal notation, and a number with prefix $0x$ is in hexadecimal notation, unless otherwise stated. The bits of a value are numbered from right to left, the leftmost bit is the most significant bit, and the rightmost bit is the least significant bit, except in the case of DES, where we use the same numbering notation as in FIPS-46 [32]. We use the following notation.

$\oplus$ bitwise logical exclusive OR (XOR) of two bit strings of the same length
$\odot$ dot product of two bit strings of the same length
$||$ string concatenation
$\lll$ left rotation of a bit string
$\circ$ functional composition. When composing functions X and Y, X $\circ$ Y denotes the function obtained by first applying X and then applying Y
e_j a 255-bit value with zeros everywhere except for bit position j, $(0 \leq j \leq 254)$
$e_{i_0,\cdots,i_j}$ the 255-bit value equal to $e_{i_0} \oplus \cdots \oplus e_{i_j}$, $(0 \leq i_0, \cdots, i_j \leq 254)$
$\mathbb{E}$ an n-bit block cipher when used with a specific user key

2.2 Differential Cryptanalysis

Differential cryptanalysis [8] takes advantage of how a specific difference in a pair of inputs of a cipher can affect a difference in the pair of outputs of the cipher, where the pair of outputs are obtained by encrypting the pair of inputs using the same key. The notion of difference can be defined in several ways; the most widely discussed is with respect to the XOR operation. The difference between the inputs is called the input difference, and the difference between the outputs of a function is called the output difference. The combination of the input difference and the output difference is called a differential. The probability of a differential is defined as follows.

Definition 1 (from [27]). *If α and β are n-bit blocks, then the probability of the differential (α, β) for $\mathbb{E}$, written $\Delta\alpha \to \Delta\beta$, is defined to be*

$$\Pr_{\mathbb{E}}(\Delta\alpha \to \Delta\beta) = \Pr_{P\in\{0,1\}^n}(\mathbb{E}(P) \oplus \mathbb{E}(P \oplus \alpha) = \beta).$$

The following result follows trivially from Definition 1:

Proposition 1 (from [27]). *If α and β are n-bit blocks, then*

$$\Pr_{\mathbb{E}}(\Delta\alpha \to \Delta\beta) = \frac{|\{x|\mathbb{E}(x) \oplus \mathbb{E}(x \oplus \alpha) = \beta, x \in \{0,1\}^n\}|}{2^n}.$$

For a random function, the expected probability of a differential for any pair (α, β) is 2^{-n}. Therefore, if $\Pr_{\mathbb{E}}(\Delta\alpha \to \Delta\beta)$ is larger than 2^{-n}, we can use the differential to distinguish $\mathbb{E}$ from a random function, given a sufficient number of chosen plaintext pairs.

Sometimes, we simply write $\Delta\alpha \xrightarrow{\mathbb{E}} \Delta\beta$ to denote the differential $\Delta\alpha \to \Delta\beta$ for $\mathbb{E}$ in this paper.

2.3 Linear Cryptanalysis

Linear cryptanalysis [29, 31] exploits correlations between a particular linear function of the input blocks and a second linear function of the output blocks. The combination of the two linear functions is called a linear approximation. The most widely used linear function involves computing the bitwise dot product operation of the block with a specific binary vector (the specific value combined with the input blocks may be different from the value applied to the output blocks). The value combined with the input blocks is called the input mask, and the value applied to the output blocks is called the output mask. The probability of a linear approximation is defined as follows.

Definition 2 (from [27]). *If α and β are n-bit blocks, then the probability of the linear approximation (α, β) for $\mathbb{E}$, written $\Gamma\alpha \to \Gamma\beta$, is defined to be*

$$\Pr_{\mathbb{E}}(\Gamma\alpha \to \Gamma\beta) = \Pr_{P\in\{0,1\}^n}(P \odot \alpha = \mathbb{E}(P) \odot \beta).$$

We refer to below the dot product $P \odot \alpha$ as the input parity, and the dot product $\mathbb{E}(P) \odot \beta$ as the output parity. The following result follows trivially from Definition 2:

Proposition 2 (from [27]). *If α and β are n-bit blocks, then*

$$\Pr_{\mathbb{E}}(\mathit{\Gamma}\alpha \to \mathit{\Gamma}\beta) = \frac{|\{x | x \odot \alpha = \mathbb{E}(x) \odot \beta, x \in \{0,1\}^n\}|}{2^n}.$$

For a random function, the expected probability of a linear approximation for any pair (α, β) is $\frac{1}{2}$. The bias of a linear approximation $\mathit{\Gamma}\alpha \to \mathit{\Gamma}\beta$, denoted by ϵ, is defined to be $\epsilon = |\Pr_{\mathbb{E}}(\mathit{\Gamma}\alpha \to \mathit{\Gamma}\beta) - \frac{1}{2}|$. Thus, if the bias ϵ is sufficiently large, we can use the linear approximation to distinguish $\mathbb{E}$ from a random function, given a sufficient number of matching plaintext-ciphertext pairs.

2.4 General Assumptions Used in Practice

Propositions 1 and 2 give the accurate probability values of a differential and a linear approximation from a theoretical point of view. However, it is usually hard to apply them in practice to a block cipher with a large block size, for example, $n = 64$ or 128 which is currently being widely used in reality, and even harder when the differential or linear approximation operates on many rounds of the cipher. In practice, for a Markov block cipher [24], a multi-round differential (or linear approximation) is usually obtained by concatenating a few one-round differential characteristics (respectively, linear approximations), and the probability of the multi-round differential (or linear approximation) is regarded as the product (respectively, the piling-up function [29]) of the probabilities of the one-round differential characteristics (respectively, linear approximations) under the following Assumption 1.

Assumption 1 *The involved round functions behave independently.*

We note that one may argue the correctness of Assumption 1 and may use a different assumption, for example, many people would like to use the assumption that the round keys are independent and uniformly distributed; however, it is not accurate, either, for generally the round keys are actually dependent, being generated from a global user key under the key schedule algorithm of the cipher. Anyway, all such assumptions require us to treat the involved rounds as independent. As mentioned in [17], this is "most often not exactly the case, but as often it is a good approximation".

Differential and linear cryptanalyses generally treat a basic unit of input (i.e. a chosen-plaintext pair for differential cryptanalysis; a known-plaintext for linear cryptanalysis) as a random variable, and assume that given a set of inputs of the basic unit, the inputs that satisfy the required property can be approximated by an independent distribution, as followed in [9, 29].

3 Differential-Linear Cryptanalysis: Previous and Our Methodologies

In this section we first review previous methodologies on differential-linear cryptanalysis, namely Langford and Hellman's and Biham et al.'s methodologies, and then give our new methodology, followed by a few implications. First observe that for simplicity we assume that the probability for a linear approximation with bias ϵ is $\frac{1}{2}+\epsilon$ in all the following descriptions; but the same results can be obtained when the probability is $\frac{1}{2}-\epsilon$.

3.1 Langford and Hellman's Methodology

In 1994 Langford and Hellman [26] introduced differential-linear cryptanalysis as a combination of differential and linear cryptanalysis, which is based on the use of a differential-linear distinguisher. To construct a differential-linear distinguisher, they treated $\mathbb{E}$ as a cascade of two sub-ciphers $\mathbb{E}_0$ and $\mathbb{E}_1$, where $\mathbb{E}=\mathbb{E}_0\circ\mathbb{E}_1$. A differential-linear distinguisher uses a (truncated) differential $\mathit{\Delta}\alpha\rightarrow\mathit{\Delta}\beta$ with probability 1 for $\mathbb{E}_0$ and a linear approximation $\mathit{\Gamma}\gamma\rightarrow\mathit{\Gamma}\delta$ with bias ϵ for $\mathbb{E}_1$, where the output difference β of the (truncated) differential has a zero value in the bit positions concerned by the input mask of the linear approximation (thus $\beta\odot\gamma=0$ holds). Let P be a plaintext chosen uniformly at random from $\{0,1\}^n$. Thus, we have $\mathbb{E}_0(P)\odot\gamma=\mathbb{E}_0(P\oplus\alpha)\odot\gamma$ with probability 1. The differential-linear distinguisher is concerned with the event $\delta\odot\mathbb{E}(P)=\delta\odot\mathbb{E}(P\oplus\alpha)$; and under Assumption 1 and the following Assumption 2 it has a probability of $\Pr(\delta\odot\mathbb{E}(P)=\delta\odot\mathbb{E}(P\oplus\alpha))=(\frac{1}{2}+\epsilon)\times(\frac{1}{2}+\epsilon)+(\frac{1}{2}-\epsilon)\times(\frac{1}{2}-\epsilon)=\frac{1}{2}+2\epsilon^2$.

Assumption 2 *The two inputs $\mathbb{E}_0(P)$ and $\mathbb{E}_0(P\oplus\alpha)$ of the linear approximation for $\mathbb{E}_1$ behave as independent inputs with respect to the linear approximation.*

Note that $\mathbb{E}(P)=\mathbb{E}_1(\mathbb{E}_0(P))$ and $\mathbb{E}(P\oplus\alpha)=\mathbb{E}_1(\mathbb{E}_0(P\oplus\alpha))$ in the above descriptions. Assumption 2 is somewhat like assuming an independent distribution for plaintext pairs generated from a particular plaintext structure with certain property in differential cryptanalysis.

By contrast, for a random function, the expected probability of a differential-linear distinguisher is $\frac{1}{2}$. Therefore, if the bias $|\Pr(\delta\odot\mathbb{E}(P)=\delta\odot\mathbb{E}(P\oplus\alpha))-\frac{1}{2}|=2\epsilon^2$ is sufficiently large, we can distinguish $\mathbb{E}$ from a random function.

3.2 Biham et al.'s Methodology

A differential-linear distinguisher plays a fundamental role in a differential-linear cryptanalysis attack. In 2002 Biham, Dunkelman and Keller [5] presented an enhanced version to make a differential-linear distinguisher cover more rounds of a block cipher, so that an attacker can potentially break more rounds of the cipher. Biham et al.'s enhanced version includes the case when the (truncated) differential $\mathit{\Delta}\alpha\rightarrow\mathit{\Delta}\beta$ has a smaller probability than 1, p say, with β meeting the condition $\beta \odot \gamma = 0$.[1] A slightly revised version was given in [14]. They applied Langford and Hellman's analysis described above when $\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) = \beta$, and used the following Assumption 3 for the cases where $\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) \neq \beta$:[2]

Assumption 3 *The output parities $\delta \odot \mathbb{E}(P)$ and $\delta \odot \mathbb{E}(P \oplus \alpha)$ have a uniform and independent distribution in $\{0,1\}$ for the cases where $\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) \neq \beta$.*

As a result, under Assumptions 1, 2 and 3, Biham et al. got $\Pr(\delta \odot \mathbb{E}(P) = \delta \odot \mathbb{E}(P \oplus \alpha)) = p \times (\frac{1}{2} + 2\epsilon^2) + (1-p) \times \frac{1}{2} = \frac{1}{2} + 2p\epsilon^2$.

Finally, they concluded that if the bias $2p\epsilon^2$ is sufficiently large, the distinguisher can be used as the basis of a differential-linear attack to distinguish $\mathbb{E}$ from a random function. Roughly, the attack has a data complexity of about $O(p^{-2}\epsilon^{-4})$.

Note. We learnt from the comments of an anonymous reviewer that the same methodology appeared earlier in Langford's PhD thesis [25], (which seems to be not publicly accessible). For simplicity, in this paper we use the phrase "Biham et al.'s methodology" to express this methodology, but hope the reader to keep in mind that Langford proposed the same methodology a few years earlier.

3.3 Our Methodology

In summary, the differential-linear distinguishers described above are concerned with the correlation between a pair of output parities, where the pair of output parities are obtained by applying a linear function (e.g. bitwise dot product with δ) to the outputs of a pair of input blocks with difference α (under the same key). The combination of the input difference and the linear function is called a differential-linear distinguisher. More formally, we define the probability of the differential-linear distinguisher as follows.

Definition 3. *If α and δ are n-bit blocks, then the probability of the differential-linear distinguisher (α, δ) for $\mathbb{E}$, written $\Delta\alpha \rightarrow \Gamma\delta$, is defined to be*

$$\Pr_{\mathbb{E}}(\Delta\alpha \rightarrow \Gamma\delta) = \Pr_{P \in \{0,1\}^n}(\mathbb{E}(P) \odot \delta = \mathbb{E}(P \oplus \alpha) \odot \delta).$$

The following result follows trivially from Definition 3:

Proposition 3. *If α and δ are n-bit blocks, then*

$$\Pr_{\mathbb{E}}(\Delta\alpha \rightarrow \Gamma\delta) = \frac{|\{x | \mathbb{E}(x) \odot \delta = \mathbb{E}(x \oplus \alpha) \odot \delta, x \in \{0,1\}^n\}|}{2^n}.$$

[1] A more general condition is $\beta \odot \gamma = c$, where $c \in \{0,1\}$ is a constant. Without loss of generality, we consider the case with $c = 0$ throughout this paper.

[2] We note that Biham et al. used a different assumption when reviewing the enhanced version in a few other papers, [7] say, where they assumed that $\mathbb{E}_0(P) \odot \gamma = \mathbb{E}_0(P \oplus \alpha) \odot \gamma$ holds with half a chance for the cases where $\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) \neq \beta$, yielding the same probability value $\frac{1}{2} + 2p\epsilon^2$ as under Assumption 3. We treat this assumption as Assumption 3, though they are different.

For a random function, the expected probability of a differential-linear distinguisher for any combination (α, δ) is $\frac{1}{2}$. Similarly, the bias of the differential-linear distinguisher $\Delta\alpha \to \Gamma\delta$ is defined to be $|\Pr_{\mathbb{E}}(\Delta\alpha \to \Gamma\delta) - \frac{1}{2}|$. Thus, if the bias is sufficiently large, we can use the differential-linear distinguisher to distinguish $\mathbb{E}$ from a random function, given a sufficient number of chosen plaintext pairs.

In practice, it is usually infeasible to compute the accurate probability of a differential-linear distinguisher $\Delta\alpha \to \Gamma\delta$ by Proposition 3, and we have to make use of some assumptions to approximate it, like Biham et al.'s methodology described in Section 3.2. However, Biham et al.'s methodology uses the three assumptions as hypotheses and works only when Assumption 3 holds; otherwise it may give probability values that are highly inaccurate in some situations; for example, let's intuitively consider the naive situation where the differential $\Delta\alpha \to \Delta\beta$ has probability $\frac{1}{2}$ and meets $\beta \odot \gamma = 0$, and all the other possible differentials $\{\Delta\alpha \to \Delta\widehat{\beta}\}$ meet $\widehat{\beta} \odot \gamma = 1$. Such an example can be easily built for a practical block cipher, DES say. The differential $\Delta\alpha \to \Delta\beta$ contributes $\frac{1}{2}[(\frac{1}{2} + \epsilon) \times (\frac{1}{2} + \epsilon) + (\frac{1}{2} - \epsilon) \times (\frac{1}{2} - \epsilon)] = \frac{1}{4} + \epsilon^2$ to the probability of the distinguisher, and the other differentials $\{\Delta\alpha \to \Delta\widehat{\beta}\}$ contribute $\frac{1}{2}[(\frac{1}{2} + \epsilon) \times (\frac{1}{2} - \epsilon) + (\frac{1}{2} - \epsilon) \times (\frac{1}{2} + \epsilon)] = \frac{1}{4} - \epsilon^2$, which also cause a bias, but in a negative way, canceling the bias due to $\Delta\alpha \to \Delta\beta$. So the real bias of the distinguisher is 0, that is, the distinguisher has no cryptanalytic significance. But if we applied Biham et al.'s methodology in this situation, the distinguisher would have a bias of $2 \times \frac{1}{2} \times \epsilon^2 = \epsilon^2$, and thus the distinguisher would be useful (if ϵ^2 is large enough); but nevertheless it is useless in fact. Notice that this case is not truly a counterexample to Biham et al.'s methodology, for it is clear that Assumption 3 does not hold for it, but it suggests that we should be cautious about using Assumption 3 and actually, we should be careful with using any assumption, and it is preferable to use as few assumptions as possible.

Biham, Dunkelman and Keller used a heuristic way to approximate the probability of a differential-linear distinguisher. We make an analysis for the probability of a differential-linear distinguisher from a mathematical point, and obtain a new methodology under only Assumptions 1 and 2. Our result is given as Theorem 1, followed by a proof.

Theorem 1. *An n-bit block cipher $\mathbb{E}$ is represented as a cascade of two subciphers $\mathbb{E}_0$ and $\mathbb{E}_1$, where $\mathbb{E} = \mathbb{E}_0 \circ \mathbb{E}_1$. If α $(\neq 0)$ is an input difference for $\mathbb{E}_0$, $\Gamma\gamma \to \Gamma\delta$ is a linear approximation with bias ϵ for $\mathbb{E}_1$, and the sum of the probabilities for the differentials $\{\Delta\alpha \to \Delta\beta | \Pr_{\mathbb{E}_0}(\Delta\alpha \to \Delta\beta) > 0, \gamma \odot \beta = 0, \beta \in \{0,1\}^n\}$ is $\widehat{p}$ $(= \sum_{\gamma \odot \beta = 0} \Pr_{\mathbb{E}_0}(\Delta\alpha \to \Delta\beta))$, then under Assumptions 1 and 2 the probability of the differential-linear distinguisher $\Delta\alpha \to \Gamma\delta$ is*

$$\Pr_{P \in \{0,1\}^n}(\mathbb{E}(P) \odot \delta = \mathbb{E}(P \oplus \alpha) \odot \delta) = \frac{1}{2} + 2(2\widehat{p} - 1)\epsilon^2.$$

Proof. Given the input difference α for $\mathbb{E}_0$, there are one or more possible output differences $\{\beta | \Pr_{\mathbb{E}_0}(\Delta\ \alpha \to \Delta\beta) > 0, \beta \in \{0,1\}^n\}$; these output differences can be classified into two sets: one is $\{\beta | \gamma \odot \beta = 0, \Pr_{\mathbb{E}_0}(\Delta\alpha \to \Delta\beta) > 0, \beta \in \{0,1\}^n\}$, and the other is $\{\beta | \gamma \odot \beta = 1, \Pr_{\mathbb{E}_0}(\Delta\alpha \to \Delta\beta) > 0, \beta \in \{0,1\}^n\}$.

Let P be a plaintext chosen uniformly at random from $\{0,1\}^n$. Then, under Assumptions 1 and 2 we have

$$\begin{aligned}
&\Pr(\mathbb{E}(P)\odot\delta=\mathbb{E}(P\oplus\alpha)\odot\delta|\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta,\gamma\odot\beta=0)\\
=&\Pr(\mathbb{E}_0(P)\odot\gamma=\mathbb{E}(P)\odot\delta,\mathbb{E}_0(P\oplus\alpha)\odot\gamma=\mathbb{E}(P\oplus\alpha)\odot\delta|\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta,\gamma\odot\beta=0)+\\
&\Pr(\mathbb{E}_0(P)\odot\gamma\neq\mathbb{E}(P)\odot\delta,\mathbb{E}_0(P\oplus\alpha)\odot\gamma\neq\mathbb{E}(P\oplus\alpha)\odot\delta|\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta,\gamma\odot\beta=0)\\
=&(\frac{1}{2}+\epsilon)\times(\frac{1}{2}+\epsilon)+[1-(\frac{1}{2}+\epsilon)]\times[1-(\frac{1}{2}+\epsilon)]\\
=&\frac{1}{2}+2\epsilon^2,
\end{aligned}$$

and

$$\begin{aligned}
&\Pr(\mathbb{E}(P)\odot\delta=\mathbb{E}(P\oplus\alpha)\odot\delta|\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta,\gamma\odot\beta=1)\\
=&\Pr(\mathbb{E}_0(P)\odot\gamma=\mathbb{E}(P)\odot\delta,\mathbb{E}_0(P\oplus\alpha)\odot\gamma\neq\mathbb{E}(P\oplus\alpha)\odot\delta|\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta,\gamma\odot\beta=1)+\\
&\Pr(\mathbb{E}_0(P)\odot\gamma\neq\mathbb{E}(P)\odot\delta,\mathbb{E}_0(P\oplus\alpha)\odot\gamma=\mathbb{E}(P\oplus\alpha)\odot\delta|\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta,\gamma\odot\beta=1)\\
=&(\frac{1}{2}+\epsilon)\times[1-(\frac{1}{2}+\epsilon)]+[1-(\frac{1}{2}+\epsilon)]\times(\frac{1}{2}+\epsilon)\\
=&\frac{1}{2}-2\epsilon^2.
\end{aligned}$$

Next, under Assumptions 1 and 2 we can compute the probability of the differential-linear distinguisher as follows.

$$\begin{aligned}
&\Pr(\mathbb{E}(P)\odot\delta=\mathbb{E}(P\oplus\alpha)\odot\delta)\\
=&\sum_{\beta\in\{0,1\}^n,Y\in\{0,1\}}\Pr(\mathbb{E}(P)\odot\delta=\mathbb{E}(P\oplus\alpha)\odot\delta,\mathbb{E}_0(P)\odot\gamma\oplus\mathbb{E}_0(P\oplus\alpha)\odot\gamma=Y,\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta)\\
=&\sum_{\beta\in\{0,1\}^n,Y\in\{0,1\}}\Pr(\mathbb{E}(P)\odot\delta=\mathbb{E}(P\oplus\alpha)\odot\delta|\mathbb{E}_0(P)\odot\gamma\oplus\mathbb{E}_0(P\oplus\alpha)\odot\gamma=Y,\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta)\times\\
&\Pr(\mathbb{E}_0(P)\odot\gamma\oplus\mathbb{E}_0(P\oplus\alpha)\odot\gamma=Y,\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta)\\
=&\sum_{\beta\in\{0,1\}^n}\Pr(\mathbb{E}(P)\odot\delta=\mathbb{E}(P\oplus\alpha)\odot\delta|\mathbb{E}_0(P)\odot\gamma\oplus\mathbb{E}_0(P\oplus\alpha)\odot\gamma=0,\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta)\times\Pr(\mathbb{E}_0(P)\odot\gamma\oplus\mathbb{E}_0(P\oplus\alpha)\odot\gamma=0,\\
&\mathbb{E}_0(P)\oplus\mathbb{E}_0(P\oplus\alpha)=\beta)+\\
&\sum_{\beta\in\{0,1\}^n}\Pr(\mathbb{E}(P)\odot\delta=\mathbb{E}(P\oplus\alpha)\odot\delta|\mathbb{E}_0(P)\odot\gamma\oplus\mathbb{E}_0(P\oplus\alpha)\odot\gamma=1,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) = \beta) \times \Pr(\mathbb{E}_0(P) \odot \gamma \oplus \mathbb{E}_0(P \oplus \alpha) \odot \gamma = 1, \\
&\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) = \beta) \qquad (1) \\
&= (\frac{1}{2} + 2\epsilon^2) \times \sum_{\beta \in \{0,1\}^n, \gamma \odot \beta = 0} \Pr(\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) = \beta) + \\
&(\frac{1}{2} - 2\epsilon^2) \times \sum_{\beta \in \{0,1\}^n, \gamma \odot \beta = 1} \Pr(\mathbb{E}_0(P) \oplus \mathbb{E}_0(P \oplus \alpha) = \beta) \\
&= \frac{1}{2} + 2(2\widehat{p} - 1)\epsilon^2. \ \square
\end{aligned}$$

Consequently, the bias of the differential-linear distinguisher $\varDelta\alpha \rightarrow \varGamma\delta$ is

$$|\Pr_{P \in \{0,1\}^n}(\mathbb{E}(P) \odot \delta = \mathbb{E}(P \oplus \alpha) \odot \delta) - \frac{1}{2}| = 2|2\widehat{p} - 1|\epsilon^2.$$

3.4 Implications

Biham et al.'s methodology requires Assumptions 1, 2 and 3, while our methodology requires only Assumptions 1 and 2. Thus, our methodology is more reasonable than Biham et al.'s methodology.

Biham et al.'s methodology holds only when Assumption 3 holds, and under the situation we have $\widehat{p} = p + (1-p)\frac{1}{2} = \frac{1}{2} + \frac{p}{2}$, meaning that the probability value obtained using Biham et al.'s methodology equals that obtained using our methodology. Thus, when Biham et al.'s methodology holds, our methodology always holds. However, our methodology holds under some situations where Biham et al.'s methodology does not hold, for example, it works for the naive situation discussed in Section 3.3 where $\widehat{p} = p = \frac{1}{2}$. Therefore, our methodology is more general than Biham et al.'s methodology. (When Langford and Hellman's methodology holds, our methodology always holds as well.)

Our methodology still requires Assumptions 1 and 2. Assumption 1 is extensively used in and is commonly regarded as necessary for differential and linear cryptanalysis in practice. Assumption 2 seems irremovable to get such a simple and practical probability formula; otherwise, the formula could not be so simple, but a more accurate version can be easily obtained from our above reasonings, for instance, from Eq. (1), though it is complicated and appears to be hardly applicable in practice. The assumptions mean that, in some cases, the probability of a differential-linear distinguisher may be overestimated or underestimated, and so is the success probability of the attack; however, computer experiments [6, 16, 23, 26, 29, 30] have shown that the assumptions work well in practice for some block ciphers. Anyway, it seems reasonable to take the worst case assumption from the point of the user of a cipher. We suggest that if possible an attacker should check the validity of these assumptions when applying them to a specific cipher.

Our result shows that using only one (truncated) differential satisfying $\beta \odot \gamma = 0$ is not sufficient in most situations, and it is likely to be not sufficient in the general situation; we should use all the differentials satisfying $\beta \odot \gamma = 0$

instead. This makes the distinguisher harder and even impossible to construct in practice, due to a large number of possible output differences. Anyway, we should use at least those differentials with a significant contribution to reduce the deviation if we are able to do so. Biham et al.'s methodology suggests that if the bias of the linear approximation keeps constant, the larger p is, the bigger is the bias of the distinguisher. Now, we know that may be not true in the general situation: A differential with a bigger probability will not necessarily result in a distinguisher with a bigger bias.

When constructing a differential-linear distinguisher, in Biham et al.'s methodology the attacker first chooses a (truncated) differential that meets the condition (as followed in [5, 6, 15, 16], in practice the output difference of the differential has zeros in the bit positions concerned by the input mask of the linear approximation), then calculates the probability of the differential, and finally takes this probability as the value of p. Our new methodology suggests a different format, that is, computing $\widehat{p}$. Once the linear approximation and the input difference of the differentials are chosen, that how many rounds can be constructed for a distinguisher depends to some extent on the computational power available for the attacker.

Our new methodology can lead to some better differential-linear cryptanalytic results than Biham et al.'s and Langford and Hellman's methodologies, as to be demonstrated by its applications to the block ciphers DES and CTC2 in the following two sections. Before further proceeding, observe that DES is a Markov cipher under the XOR difference notion [24], and similarly we can learn that CTC2 as well as Serpent is a Markov cipher under the XOR difference notion.

At last, to be conservative, we would like to suggest that one should pay attention to all these methodologies, for a real situation is usually hard to predict, and it may make the Assumption 3 for Biham et al.'s methodology hold.

4 Application to the DES Block Cipher

The DES block cipher is well known to both academia and industry, which has a 64-bit block size, a 56-bit user key, and a total of 16 rounds. We refer the reader to [32] for the specifications of DES.

In 1994, under the two default Assumptions 1 and 2 Langford and Hellman [26] used their methodology to obtain a 6-round differential-linear distinguisher of DES, and finally applied it to break 8-round DES; the attack recovers 16 key bits with a time complexity of $2^{14.6}$ 8-round DES encryptions, so it would take 2^{40} encryptions to recover the remaining 40 key bits with an exhaustive search, meaning that a total of approximately 2^{40} 8-round DES encryptions are required to recover the whole 56 key bits (Note that there might exist an efficient way to obtain the remaining key bits). In 2002, under Assumptions 1, 2 and 3, Biham, Dunkelman and Keller [5] described a 7-round differential-linear distinguisher of DES using their enhanced methodology, and finally gave differential-linear attacks on 8 and 9-round DES; and an improved version of the 9-round attack appeared in pages 108–111 of [14]. Their attack recovers 18 key bits with a time complexity of $2^{29.17}$ 9-round DES encryptions, the remaining 38 key bits would take 2^{38} encryptions to recover with a key exhaustion, and thus it has a total of approximately 2^{38} 9-round DES encryptions to recover the whole 56 key bits.

Nevertheless, we find that our new methodology enables us to construct 7 and 8-round differential-linear distinguishers of DES based on the same 3-round linear approximation as used in the previous differential-linear cryptanalysis of DES [5,26]; the 8-round distinguisher can allow us to break 10-round DES. More importantly, we are able to construct a 11-round differential-linear distinguisher of DES, and finally use it as the basis of a differential-linear attack on 13-round DES. Below we describe the 11-round differential-linear distinguisher and our attack on 13-round DES. We write the subkey used in the S_l S-box of Round m as $K_{m,l}$, where $1 \leq m \leq 16, 1 \leq l \leq 8$.

4.1 A 11-Round Differential-Linear Distinguisher with Bias $2^{-24.05}$

The 11-round differential-linear distinguisher is made up of a 6-round linear approximation $\mathit{\Gamma}\gamma \rightarrow \mathit{\Gamma}\delta$ with bias $1.95 \times 2^{-9} \approx 2^{-8.04}$ and all the 5-round differentials $\{\mathit{\Delta}\alpha \rightarrow \mathit{\Delta}\beta\}$ with $\mathit{\Delta}\alpha = 0x4000000000000000$. The 6-round linear approximation $\mathit{\Gamma}\gamma \rightarrow \mathit{\Gamma}\delta$ is $0x0000000001040080 \rightarrow 0x2104008000008000$, (which is the best 6-round linear approximation given in [29]). Let's compute the probability of the 11-round differential-linear distinguisher using our new methodology.

We first consider the 5-round differentials $\{\mathit{\Delta}\alpha \rightarrow \mathit{\Delta}\beta\}$. There is a one probability in the first round, meaning that the first round is bypassed by the differential characteristic with probability 1. After the **E** expansion operation of the second round, $0x4$ in $\mathit{\Delta}\alpha$ becomes $0x8$, which enters the S_1 S-box of the second round and generates 11 differences after the S-box: $\{\omega|\omega = 0x3, 0x5, 0x6, 0x7, 0x9, 0xA, 0xB, 0xC, 0xD, 0xE, 0xF\}$; the probabilities for the output differences are given in the second column of Table 2. We represent ω as a concatenation of four one-bit variables $a||b||c||d$, where $a, b, c, d \in \{0, 1\}$. Thus, the right half of the third round has the input difference $00000000a0000000b00000 c0000000d0$ in binary notation, and this input difference can make at most 6 S-boxes of the third round active: $\mathrm{S}_2, \mathrm{S}_3, \mathrm{S}_4, \mathrm{S}_5, \mathrm{S}_6, \mathrm{S}_8$.

In the third round, the S_2 S-box has an input difference $00000a$ in binary notation, the S_3 S-box has an input difference $0a0000$ in binary notation, the S_4 S-box has an input difference $00000b$ in binary notation, the S_5 S-box has an input difference $0b0000$ in binary notation, the S_6 S-box has an input difference $000c00$ in binary notation, and the S_8 S-box has an input difference $000d00$ in binary notation. We denote respectively by x_0, x_1, x_2 the most significant bit, the second most significant bit and the second least significant bit of the output difference of the S_2 S-box, by $x_3||x_4||x_5||x_6$ the output difference of the S_3 S-box, by x_7, x_8, x_9 the second most significant bit, the second least significant bit and the least significant bit of the output difference of the S_4 S-box, by $x_{10}||x_{11}||x_{12}||x_{13}$ the output difference of the S_5 S-box, by x_{14}, x_{15}, x_{16} the most

Table 2. Probabilities for the eleven output differences in $\{\omega\}$

ω	$\mathrm{Pr}_{\mathrm{S}_1}(\varDelta 0x8 \rightarrow \varDelta\omega)$	$\mathrm{Pr}(\varDelta\beta_\omega \odot \varGamma\gamma = 0 \mid \varDelta 0x8 \rightarrow \varDelta\omega)$
$0x3$	$\frac{12}{64}$	0.49779944866895676
$0x5$	$\frac{8}{64}$	0.49595199525356293
$0x6$	$\frac{8}{64}$	0.50433863041689619
$0x7$	$\frac{4}{64}$	0.50256029706542904
$0x9$	$\frac{6}{64}$	0.50855094581311278
$0xA$	$\frac{2}{64}$	0.50591027818154544
$0xB$	$\frac{8}{64}$	0.50239421910760029
$0xC$	$\frac{8}{64}$	0.49929085310759547
$0xD$	$\frac{2}{64}$	0.49968796220765910
$0xE$	$\frac{2}{64}$	0.50061782109781916
$0xF$	$\frac{4}{64}$	0.50005227406592345

significant bit, the second most significant bit and the second least significant bit of the output difference of the S_6 S-box, and by x_{17}, x_{18}, x_{19} the most significant bit, the second least significant bit and the least significant bit of the output difference of the S_8 S-box.

In the fourth round, the S_1 S-box has the input difference $0||x_9||(x_2 \oplus 1)||x_{13}||x_{14}||x_{17}$, and we denote by y_0 the second most significant bit of its output difference; the S_2 S-box has the input difference $x_{14}||x_{17}||x_6||0||x_{10}||0$, and we denote by y_1 the least significant bit of its output difference; the S_3 S-box has the input difference $x_{10}||0||x_8||x_{16}||0||x_0$, and we denote by y_2 the second most significant bit of its output difference; the S_4 S-box has the input difference $0||x_0||x_{11}||x_{18}||x_4||0$, and we denote by y_3 the second most significant bit of its output difference; the S_6 S-box has the input difference $x_7||x_{19}||0||0||x_3||x_{12}$, and we denote by y_4 the least significant bit of its output difference; the S_8 S-box has the input difference $x_1||x_{15}||x_5||0||0||x_9$, and we denote by y_5 the least significant bit of its output difference. Thus we have that the input difference of the S_5 S-box of the fifth round is $y_2||(y_0 \oplus b)||y_1||y_4||y_3||y_5$.

A simple analysis reveals that the three bits concerned by the input mask $\varGamma\gamma$ depend on: (1) x_{10}, x_{11} and x_{12}; and (2) The three most significant bits of the output difference of the S_5 S-box of the fifth round; and we denote the XOR of the three bits by z.

For each difference ω, we denote by β_ω the output difference(s) of the 5-round DES. Now, by the differential distribution tables of the S-boxes (see [9]) we can compute the probability that the XOR of the concerned three bits of β_ω (i.e., $x_{10} \oplus x_{11} \oplus x_{12} \oplus z$) is zero by performing a computer program over all the possible (truncated) differential characteristics. These probabilities are given in the third column of Table 2. The largest number of possible differential characteristics happens when $\omega = 0xF$, which is $7 \times 10 \times 4 \times 10 \times 6 \times 7 \times 2^6 \times 2 \approx 2^{23.9}$; and it takes a few seconds to check on a personal computer.

Finally, by Theorem 1 we have that the probability of the 11-round distinguisher $\varDelta\alpha \rightarrow \varGamma\delta$ is $\frac{1}{2} + 2 \times [2 \times \sum_{\omega} \Pr_{\mathrm{S}_1}(\varDelta 0x8 \rightarrow \varDelta\omega) \times \Pr(\varDelta\beta_{\omega} \odot \varGamma\gamma = 0|\varDelta 0x8 \rightarrow \varDelta\omega) - 1] \times (2^{-8.04})^2 \approx \frac{1}{2} + 2 \times 2^{-8.97} \times (2^{-8.04})^2 = \frac{1}{2} + 2^{-24.05}$. Therefore, the 11-round distinguisher has a bias of $2^{-24.05}$.

4.2 Differential-Linear Attack on 13-Round DES

The 11-round distinguisher $\varDelta\alpha \rightarrow \varGamma\delta$ can be used to break 13-round DES. We assume the attacked rounds are the first thirteen rounds from Rounds 1 to 13. A simple analysis on the key schedule of DES reveals that $K_{1,1}$ and $K_{13,1}$ overlap in 2 bits (i.e. bits 17 and 34 of the user key), and thus given $K_{1,1}$ we know 2 bits of $K_{13,1}$. The attack procedure is as follows.

1. Choose $2^{47.1}$ structures $\mathcal{S}_i$, $(i = 1, 2, \cdots, 2^{47.1})$, where a structure is defined to be a set of 2^4 plaintexts $P_{i,j}$ with bits (9,17,23, 31) of the left half taking all the possible values, bit (2) of the right half fixed to 0 and the other 59 bits fixed, $(j = 1, 2, \cdots, 2^4)$. In a chosen-plaintext attack scenario, obtain all the ciphertexts for the 2^4 plaintexts in each of the $2^{47.1}$ structures; we denote by $C_{i,j}$ the ciphertext for plaintext $P_{i,j}$.
2. Choose $2^{47.1}$ structures $\widehat{\mathcal{S}}_i$, $(i = 1, \cdots, 2^{47.1})$, where a structure $\widehat{\mathcal{S}}_i$ is obtained by setting 1 to bit (2) of the right half of all the plaintexts $P_{i,j}$ in $\mathcal{S}_i$. In a chosen-plaintext attack scenario, obtain all the ciphertexts for the 2^4 plaintexts in each $\widehat{\mathcal{S}}_i$.
3. Guess a value for $K_{1,1}$, and do as follows.
   (a) Initialize 2^{20} counters to zero, which correspond to the 2^{20} possible pairs consisting of the possible values for a couple of the 10 ciphertext bits: bit (17) of the left half and bits (1,2,3,4,5,8,14,25,32) of the right half.
   (b) Partially encrypt every (remaining) plaintext $P_{i,j}$ with the guessed $K_{1,1}$ to get its intermediate value immediately after Round 1; we denote it by $\varepsilon_{i,j}$.
   (c) Partially decrypt $\varepsilon_{i,j} \oplus 0x4000000000000000$ with the guessed $K_{1,1}$ to get its plaintext, and find the plaintext in $\widehat{\mathcal{S}}_i$; we denote it by $\widehat{P}_{i,j}$, and denote by $\widehat{C}_{i,j}$ the corresponding ciphertext for $\widehat{P}_{i,j}$. Store $(C_{i,j}, \widehat{C}_{i,j})$ in a table.
   (d) For every ciphertext pair $(C_{i,j}, \widehat{C}_{i,j})$, add 1 to the counter corresponding to the pair of the 10 ciphertext bits specified by $(C_{i,j}, \widehat{C}_{i,j})$.
   (e) Guess a value for the unknown 4 bits of $K_{13,1}$, and do as follows.
      i. For each of the 2^{20} pairs of the concerned 10 ciphertext bits, partially decrypt it with the guessed $K_{13,1}$ to get the pair of the 5 bits concerned by the output mask $\varGamma\delta$, and compute the XOR of the pair of the 5 bits (concerned by the output mask).
      ii. Count the number of the ciphertext pairs $(C_{i,j}, \widehat{C}_{i,j})$ such that the XOR of the pair of the 5 bits concerned by $\varGamma\delta$ is zero, and compute its deviation from $2^{50.1}$.

iii. If the guess for $(K_{1,1}, K_{13,1})$ is the first guess for $(K_{1,1}, K_{13,1})$, then record the guess and the deviation computed in Step 3(e)(ii); otherwise, record the guess and its deviation only when the deviation is larger than that of the previously recorded guess, and remove the guess with the smaller deviation.

4. For the $(K_{1,1}, K_{13,1})$ recorded in Step 3(e)(iii), exhaustively search for the remaining 46 key bits with two known plaintext-ciphertext pairs. If a 56-bit key is suggested, output it as the user key of the 13-round DES.

The attack requires $2^{52.1}$ chosen plaintexts. The required memory for the attack is dominated by the storage of the plaintexts and ciphertexts, which is $2^{52.1} \times 16 = 2^{56.1}$ bytes. Steps 1 and 2 have a time complexity of $2^{52.1}$ 13-round DES encryptions. Steps 3(b) and 3(c) have a time complexity of $2 \times 2^{51.1} \times 2^6 \times \frac{1}{8\times13} \approx 2^{51.4}$ 13-round DES encryptions. Step 3(d) has a time complexity of $2^{51.1} \times 2^6 = 2^{57.1}$ memory accesses. Roughly, an extremely conservative estimate is: 13 memory accesses equal a 13-round DES encryption in terms of time, assuming that the 13-round DES is implemented with 8 parallel S-box lookups per round and one round is equivalent to one memory access. So the time complexity of Step 3(d) is equivalent to $\frac{2^{57.1}}{13} \approx 2^{53.4}$ 13-round DES encryptions. The time complexity of Step 3(e) is dominated by the time complexity of Step 3(e)(i), which is $2 \times 2^6 \times 2^4 \times 2^{20} \times \frac{1}{8\times13} \approx 2^{24.3}$ 13-round DES encryptions. Step 4 has a time complexity of 2^{46} 13-round DES encryptions. Therefore, the attack has a total time complexity of approximately $2^{54.2}$ 13-round DES encryptions, faster than exhaustive key search. There are $2^{51.1}$ plaintext pairs $(P_{i,j}, \widehat{P}_{i,j})$ for a guess of $(K_{1,1}, K_{13,1})$, and thus following Theorem 2 of [33], we can know that the attack has a success probability of about 99%.

This shows that our new methodology enables us to break more rounds of DES than Biham et al.'s or Langford and Hellman's methodology. Since our attack works under only two assumptions, it is more reasonable than Biham et al.'s attack.

Note. Using the new methodology we can obtain a few differential-linear distinguishers operating on a smaller number of rounds, for example, a 7-round distinguisher $(\varDelta\alpha = 0x4000000000000000, \varGamma\delta = 0x2104008000008000)$ with bias $2^{-7.94}$ and an 8-round distinguisher $(\varDelta\alpha = 0x4000000000000000, \varGamma\delta = 0x2104008000008000)$ with bias $2^{-12.83}$, both using the same 3-round linear approximation as used in Biham et al.'s and Langford and Hellman's differential-linear cryptanalysis of DES. These distinguishers can allow us to break DES with a smaller number of rounds at a smaller complexity, for example, the 8-round distinguisher can similarly be used to break 10-round DES with a data complexity of $2^{29.66}$ chosen plaintexts and a time complexity of 2^{44} 10-round DES encryptions at a success rate of about 97%.

5 Application to the CTC2 Block Cipher

The CTC2 [11] cipher was designed to show the strength of algebraic cryptanalysis [12] on block ciphers by the proposer of algebraic cryptanalysis, who described an algebraic attack on 6 rounds of the version of CTC2 that uses a 255-bit block size and a 255-bit key. Using Biham et al.'s methodology, in 2009 Dunkelman and Keller [15] described 6 and 7-round differential-linear distinguishers for the version of CTC2, and finally presented differential-linear attacks on 7 and 8 rounds of CTC2 (with a 255-bit block size and key). The 8-round attack is known as the best previously published cryptanalytic result on the version of CTC2 in terms of the numbers of attacked rounds.

In this section, we first describe a flaw in the previous differential-linear cryptanalysis of CTC2. Then, under the new methodology we present an 8.5-round differential-linear distinguisher with bias 2^{-68} for the CTC2 with a 255-bit block size and key, and finally give a differential-linear attack on 10-round CTC2 (with a 255-bit block size and a key). First we briefly describe the CTC2 cipher.

5.1 The CTC2 Block Cipher

The CTC2 [11] block cipher has a variable block size, a variable length key and a variable number of rounds. There are many combinations for the block size, key size and round number. As in [15], we only consider the version of CTC2 that uses a 255-bit block size and a 255-bit key. CTC2 uses the following two elementary operations to construct its round function.

- **S** is a non-linear substitution operation constructed by applying the same 3×3-bit bijective S-box 85 times in parallel to an input.
- **D** is a linear diffusion operation, which takes a 255-bit block $Y = (Y_{254}, \cdots, Y_1, Y_0)$ as input, and outputs a 255-bit block $Z = (Z_{254}, \cdots, Z_1, Z_0)$, computed as defined below.

$$\begin{cases} Z_{151} = Y_2 \oplus Y_{139} \oplus Y_{21} \\ Z_{(i\times 202+2) \bmod 255} = Y_i \oplus Y_{(i+137) \bmod 255} \qquad i = 0, 1, 3, 4, \cdots, 254 \end{cases}$$

CTC2 takes as input a 255-bit plaintext block P, and its encryption procedure for N_r rounds is, where Z_0, X_i, Y_i, Z_i, $X_{N_r}, Y_{N_r}, Z_{N_r}$ are 255-bit variables, and K_0, K_i, K_{N_r} are round keys generated from a user key K as $K_j = K \lll j$ in our notation, $(0 \leq j \leq N_r)$.

```
1. $Z_0 = P$.
2. For $i = 1$ to $N_r - 1$:
     – $X_i = Z_{i-1} \oplus K_{i-1}$,
     – $Y_i = \mathbf{S}(X_i)$,
     – $Z_i = \mathbf{D}(Y_i)$.
3. $X_{N_r} = Z_{N_r-1} \oplus K_{N_r-1}$, $Y_{N_r} = \mathbf{S}(X_i)$, $Z_{N_r} = \mathbf{D}(Y_{N_r})$.
4. Ciphertext $= Z_{N_r} \oplus K_{N_r}$.
```

To keep in accordance with [11], the ith iteration of Step 2 in the above description is referred to as Round i, $(1 \leq i \leq N_r - 1)$, and the transformations in Steps 3 and 4 are referred to as Round N_r. We number the 85 S-boxes in a round from 0 to 84 from right to left.

5.2 A Flaw in Previous Differential-Linear Cryptanalysis of CTC2

Observe that Dunkelman and Keller used the 0.5-round differential $e_{30,151} \overset{\mathbf{D}}{\to} e_2$ with probability 1 in their differential-linear attacks presented in [15]. However, we find that this differential is not correct: For the $\mathbf{D}$ operation, given the input difference $e_{30,151}$, we cannot get the output difference e_2; and the correct output difference should be $e_{25,63,159,197}$. On the other hand, for the $\mathbf{D}$ operation, given the output difference e_2, the input difference has over fifty non-zero bits, much more than the number two in $e_{30,151}$. As a consequence, the differential-linear cryptanalytic results are flawed.

Note that Dunkelman and Keller also described differential attacks on 5, 6 and 7-round CTC2 in [15], and the 0.5-round differential $e_{30,151} \overset{\mathbf{D}}{\to} e_2$ with probability 1 was also used and played a very important role in the differential results; thus they are flawed, too. It seems very hard to correct these differential and differential-linear cryptanalytic results to break that many rounds of CTC2.

5.3 An 8.5-Round Differential-Linear Distinguisher with Bias 2^{-68}

The 8.5-round differential-linear distinguisher with bias 2^{-68} is made up of a 5.5-round linear expression $\Gamma\gamma \to \Gamma\delta$ with bias 2^{-33} and all the 3-round differentials $\{\Delta\alpha \to \Delta\beta\}$ with $\Delta\alpha = e_0$. The 5.5-round linear expression $\Gamma\gamma \to \Gamma\delta$ is $e_{5,33,49,54,101,112,131,138,155,168,188,193,217,247,251} \to e_{32,151}$. Using the new methodology we can compute that the 8.5-round distinguisher $\Delta\alpha \to \Gamma\delta$ has a bias of 2^{-68}, in a manner similar to that for the above 11-round DES distinguisher.

5.4 Differential-Linear Attack on 10-Round CTC2 with a 255-Bit Block Size and Key

The above 8.5-round distinguisher can be used as the basis for a differential-linear attack breaking the version of CTC2 that has a 255-bit block size, a 255-bit key and a total of 10 rounds.

We assume the attacked rounds are the first ten rounds from Rounds 1 to 10; and we use the distinguisher from Rounds 2 until before the $\mathbf{D}$ operation of Round 10. We can learn that the input difference α propagates to 16 bit positions after the inverse of the $\mathbf{D}$ operation of Round 1: Bits 17, 21, 40, 59, 78, 97, 116, 135, 139, 154, 158, 177, 196, 215, 234 and 253. The 16 active bits correspond to 16 S-boxes of Round 0: S-boxes 5, 7, 13, 19, 26, 32, 38, 45, 46, 51, 52, 59, 65, 71, 78 and 84; let Θ be the set of the 16 S-boxes, and K_Θ be the 48 bits of K_0 corresponding to the 16 S-boxes in Θ. Another observation is that we do not need to guess the subkey bits from K_{10}, because the output mask $\Gamma\delta$ of the 8.5-round distinguisher concerns the intermediate value immediately after the $\mathbf{S}$ operation of Round 10, and for a pair of ciphertexts $(C, \widehat{C})$ the value of $\delta \odot \mathbf{D}^{-1}(C) \oplus \delta \odot \mathbf{D}^{-1}(\widehat{C})$ equals to $\delta \odot \mathbf{D}^{-1}(C \oplus \widehat{C})$, which is independent of K_{10}. The attack procedure is as follows.

1. Choose 2^{94} structures $\mathcal{S}_i$, $(i = 0, 1, \cdots, 2^{94} - 1)$, where a structure is defined to be a set of 2^{48} plaintexts $P_{i,j}$ with the 48 bits for the S-boxes in Θ taking all the possible values and the other 207 bits fixed, $(j = 0, 1, \cdots, 2^{48} - 1)$. In a chosen-plaintext attack scenario, obtain all the ciphertexts for the 2^{48} plaintexts in each of the 2^{94} structures; we denote by $C_{i,j}$ the ciphertext for plaintext $P_{i,j}$.
2. Initialize 2^{48} counters to zero, which correspond to all the possible values for K_Θ.
3. For every structure $\mathcal{S}_i$, guess a value for K_Θ, and do as follows.
   (a) Partially encrypt every (remaining) plaintext $P_{i,j}$ with the guessed K_Θ to get its intermediate value immediately after the **S** operation of Round 1; we denote it by $\varepsilon_{i,j}$.
   (b) Take bitwise complements to bits (17, 21, 40, 59, 78, 97, 116, 135, 139, 154, 158, 177, 196, 215, 234, 253) of $\varepsilon_{i,j}$, and keep the other bits of $\varepsilon_{i,j}$ invariant; we denote the resulting value by $\widehat{\varepsilon}_{i,j}$.
   (c) Partially decrypt $\widehat{\varepsilon}_{i,j}$ with the guessed K_Θ to get its plaintext, and find the plaintext in $\mathcal{S}_i$; we denote it by $\widehat{P}_{i,j}$, and denote by $\widehat{C}_{i,j}$ the corresponding ciphertext for $\widehat{P}_{i,j}$.
   (d) For $(C_{i,j}, \widehat{C}_{i,j})$, compute the XOR of bits 32 and 151 of $\mathbf{D}^{-1}(C_{i,j} \oplus \widehat{C}_{i,j})$. If the XOR is zero, add 1 to the counter corresponding to the guessed K_Θ.
4. For the K_Θ with the highest deviation from 2^{140}, exhaustively search for the remaining 207 key bits with a known plaintext-ciphertext pair. If a 255-bit key is suggested, output it as the user key of the version of CTC2.

The attack requires 2^{142} chosen plaintexts. Note that we start to collect another structure of plaintexts only after testing a structure of plaintexts, so that we can reuse the memory for storing the structure of plaintexts, hence the required memory of the attack is dominated by the storage of the 2^{48} counters and a structure of 2^{48} plaintext-ciphertext pairs, which is $2^{48} \times \frac{48}{8} + 2 \times 2^{48} \times \frac{255}{8} \approx 2^{54.2}$ bytes of memory. The time complexity of Step 3 is dominated by the time complexity of Steps 3(a), 3(c) and 3(d), which is approximately $2 \times 2^{141} \times 2^{48} \times \frac{16}{85 \times 10} + 2^{141} \times 2^{48} \times \frac{1}{10} \approx 2^{186.2}$ 10-round CTC2 encryptions. Step 4 has a time complexity of 2^{207} 10-round CTC2 encryptions. Therefore, the attack has a total time complexity of 2^{207} 10-round CTC2 encryptions to find the 255-bit key. There are 2^{141} plaintext pairs $(P_{i,j}, \widehat{P}_{i,j})$ for a guess of K_Θ. Following Theorem 2 of [33], we can learn that the probability that the correct guess for K_Θ has the highest deviation is about 99.9%. Thus, the attack has a success probability of about 99.9%.

6 Possible Extensions of Our Methodology

In this section we briefly discuss several possible extensions of our methodology, although particulars should be noticed.

The first possible extension is to consider the case when using two different values for the output mask δ in Definition 3, say δ_1, δ_2; that is, we might consider the event $\mathbb{E}(P) \odot \delta_1 = \mathbb{E}(P \oplus \alpha) \odot \delta_2$ for a randomly chosen $P \in \{0,1\}^n$. The resulting differential-linear distinguisher would have a bias of $2(2\widehat{p}-1)\epsilon_1\epsilon_2$ for some ϵ_1 and ϵ_2 denoting the respective bias of the two linear approximations. From a theoretical point of view, there seems no need to use two different output masks, for we can always choose the output mask with a bigger bias, and a key-recovery attack based on a differential-linear distinguisher with two different output masks requires us to guess no less key bits than that based on a differential-linear distinguisher with one output mask; however, the case with two different output masks may depend on Assumption 2 to a lesser degree than the discussed case with one output mask, for the two linear approximations can be independent somewhat, instead of two identical linear approximations used in the case with one output mask, and thus it may potentially be particularly helpful when making a practicable attack in reality.

The second possible extension is to consider the case when applying our methodology in a related-key [3,19,21] attack scenario. The notion of the related-key differential-linear analysis appeared in [18], and later Kim [20] described an enhanced version based on Biham et al.'s enhanced methodology. Likewise, we can get a more reasonable and general version based on our new methodology.

Other possible extensions are to obtain new methodologies, in a way similar to the above new methodology for differential-linear cryptanalysis, for the high-order differential-linear attack, the differential-bilinear attack and the differential-bilinear-boomerang attack, which were proposed in [7]. At present, however, these attack techniques appear to be hard to apply to obtain good cryptanalytic results in practice.

7 Conclusions

In this paper we have given a new methodology for differential-linear cryptanalysis under only the two assumptions implicitly used in the very first published paper on this technique. The new methodology is more reasonable and more general than Biham et al.'s methodology, and it can lead to some better differential-linear cryptanalytic results for some block ciphers than the previously known methodologies.

Using the new methodology, we have presented differential-linear attacks on 13-round DES and 10-round CTC2 with a 255-bit block size and key. In terms of the numbers of attacked rounds, the 10-round CTC2 attack is the first published cryptanalytic attack on the version of CTC2; and the 13-round DES attack is much better than any previously published differential-linear cryptanalytic results for DES, though it is inferior to the best previously published cryptanalytic results for DES. In addition, an important merit for these new differential-linear cryptanalytic results is that they are obtained under only two assumptions and thus are more reasonable than those obtained using Biham et al.'s methodology. Like most cryptanalytic results on block ciphers, most of these attacks are far less than practical at present, but they provide a comprehensive understanding of the security of the block ciphers.

The new methodology is a general cryptanalysis technique and can be potentially used to cryptanalyse other block ciphers; and block cipher designers should pay attention to this new methodology when designing ciphers.

The new methodology still requires Assumptions 1 and 2. As a direction for future research on differential-linear cryptanalysis, it would be interesting to investigate how to further reduce the number of assumptions used, making a more reasonable and more general methodology that could be used in practice.

Acknowledgments

The author is very grateful to Dr. Orr Dunkelman and Dr. Nathan Keller for their discussions on the flaw about CTC2 and to the anonymous referees for their comments on earlier versions of the paper.

References

1. Anderson, R., Biham, E., Knudsen, L.R.: Serpent: a new block cipher proposal. In: Vaudenay, S. (ed) FSE 1998. LNCS, vol. 1372, pp. 222–238. Springer, Heidelberg (1998).
2. Anderson, R., Biham, E., Knudsen, L.R.: Serpent: a proposal for the Advanced Encryption Standard, 1998.
3. Biham, E.: New types of cryptanalytic attacks using related keys. Journal of Cryptology 7(4), 229–246 (1994)
4. Biham, E., Biryukov, A.: An improvement of Davies' attack on DES. Journal of Cryptology 10(3), 195–206 (1997)
5. Biham, E., Dunkelman, O., Keller, N.: Enhancing differential-linear cryptanalysis. In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 254–266. Springer, Heidelberg (2002)
6. Biham, E., Dunkelman, O., Keller, N.: Differential-linear cryptanalysis of Serpent. In: Johansson, T. (ed.) FSE 2003. LNCS, vol. 2887, pp. 9–21. Springer, Heidelberg (2003)
7. Biham, E., Dunkelman, O., Keller, N.: New combined attacks on block ciphers. In: Gilbert, H., Handschuh, H. (eds.) FSE 2005. LNCS, vol. 3557, pp. 126–144. Springer, Heidelberg (2005)
8. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. In: Menezes, A., Vanstone, S.A. (eds.) CRYPTO 1990. LNCS, vol. 537, pp. 2–21. Springer, Heidelberg (1990)
9. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. Journal of Cryptology 4(1), 3–72 (1991)
10. Biham, E., Shamir, A.: Differential cryptanalysis of the full 16-round DES. In: Brickell, E.F. (ed.) CRYPTO 1992. LNCS, vol. 740, pp. 487–496. Springer, Heidelberg (1993)
11. Courtois, N.T.: CTC2 and fast algebraic attacks on block ciphers revisited. IACR ePrint report 2007/152 (2007)

12. Courtois, N.T., Pieprzyk, J.: Cryptanalysis of block ciphers with overdefined systems of equations. In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 267–287. Springer, Heidelberg (2002)
13. Davies, D.: Investigation of a potential weakness in the DES algorithm. (1987)
14. Dunkelman, O.: Techniques for cryptanalysis of block ciphers. Ph.D. thesis, Technion — Israel Institute of Technology, Israel (2006)
15. Dunkelman, O., Keller, N.: Cryptanalysis of CTC2. In: Fischlin, M. (ed.) CT-RSA 2009. LNCS, vol. 5473, pp. 226–239. Springer, Heidelberg (2009)
16. Dunkelman, O., Indesteege, S., Keller, N.: A differential-linear attack on 12-round Serpent. In: Chowdhury, D.R., Rijmen, V., Das, A. (eds.) INDOCRYPT 2008. LNCS, vol. 5365, pp. 308–321. Springer, Heidelberg (2008)
17. Handschuh, H., Naccache, D.: SHACAL. In: Proceedings of the First Open NESSIE Workshop (2000)
18. Hawkes, P.: Differential-linear weak key classes of IDEA. In: Nyberg, K. (ed.) EUROCRYPT 1998. LNCS, vol. 1403, pp. 112–126. Springer, Heidelberg (1998)
19. Kelsey, J., Schneier, B., Wagner, D.: Key-schedule cryptanalysis of IDEA, G-DES, GOST, SAFER, and Triple-DES. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 237–251. Springer, Heidelberg (1996)
20. Kim, J.: Combined differential, linear and related-key attacks on block ciphers and MAC algorithms. Ph.D. thesis, Katholieke Universiteit Leuven, Blegium (2006)
21. Knudsen, L.R.: Cryptanalysis of LOKI91. In: Seberry, J., Zheng, Y. (eds.) ASIACRYPT 1992. LNCS, vol. 718, pp. 196–208. Springer, Heidelberg (1993)
22. Knudsen, L.R.: Trucated and higher order differentials. In: Preneel, B. (ed.) FSE 1994. LNCS, vol. 1008, pp. 196–211. Springer, Heidelberg (1995)
23. Knudsen, L.R., Mathiassen, J.E.: A chosen-plaintext linear attack on DES. In: Schneier, B. (ed.) FSE 2000. LNCS, vol. 1978, pp. 262–272. Springer, Heidelberg (2001)
24. Lai, X., Massey, J.L., Murphy, S.: Markov ciphers and differential cryptanalysis. In: Davies, D.W. (ed.) EUROCRYPT 1991. LNCS, vol. 547, pp. 17–38. Springer, Heidelberg (1991)
25. Langford, S.K.: Differential-linear cryptanalysis and threshold signatures. Ph.D. thesis, Stanford University, USA (1995)
26. Langford, S.K., Hellman, M.E.: Differential-linear cryptanalysis. In: Desmedt, Y. (ed.) CRYPTO 1994. LNCS, vol. 839, pp. 17–25. Springer, Heidelberg (1994)
27. Lu, J.: Cryptanalysis of block ciphers. Ph.D. thesis, University of London, UK (2008)
28. Lu, J.: New methodologies for differential-linear cryptanalysis and its extensions. Cryptology ePrint Archive, Report 2010/025 (2010). `http://eprint.iacr.org/2010/025`
29. Matsui, M.: Linear cryptanalysis method for DES cipher. In: Helleseth, T. (ed.) EUROCRYPT 1993. LNCS, vol. 765, pp. 386–397. Springer, Heidelberg (1994)
30. Matsui, M.: The first experimental cryptanalysis of the Data Encryption Standard. In: Desmedt, Y.G. (ed.) CRYPTO 1994. LNCS, vol. 839, pp. 1–11. Springer, Heidelberg (1994)
31. Matsui, M., Yamagishi, A.: A new method for known plaintext attack of FEAL cipher. In: Rueppel, R.A. (ed.) EUROCRYPT 1992. LNCS, vol. 658, pp. 81–91. Springer, Heidelberg (1993)
32. National Bureau of Standards (NBS), Data Encryption Standard (DES), FIPS-46 (1977)
33. Selçuk, A.A.: On probability of success in linear and differential cryptanalysis. Journal of Cryptology 21(1), 131–147 (2008)