Linear Cryptanalysis Using Multiple Approximations

Burton S. Kaliski Jr. and M.J.B. Robshaw

RSA Laboratories
100 Marine Parkway
Redwood City, CA 94065, USA

Abstract. We present a technique which aids in the linear cryptanalysis of a block cipher and allows for a reduction in the amount of data required for a successful attack. We note the limits of this extension when applied to DES, but illustrate that it is generally applicable and might be exceptionally successful when applied to other block ciphers. This forces us to reconsider some of the initial attempts to quantify the resistance of block ciphers to linear cryptanalysis, and by taking account of this new technique we cover several issues which have not yet been considered.

1 Introduction

Matsui and Yamagishi [6] introduced the idea of *linear cryptanalysis* in 1992 in an attack on FEAL [10]. The techniques used in this attack were refined by Matsui and used with dramatic effect on DES [7] in a theoretical attack on the full 16-round DES requiring 2^{47} known plaintext/ciphertext pairs [4]. After further work an experiment was performed during which the key used in the full 16-round version of DES was recovered using 2^{43} known plaintext/ciphertext pairs [9].

The most notable feature about linear cryptanalysis is that it is a known plaintext attack rather than a chosen plaintext attack (differential cryptanalysis [1] is a chosen plaintext attack) and as such poses more of a practical threat to a block cipher. At present, however, a successful linear cryptanalytic attack on DES still requires a large quantity of known plaintext.

In this paper we consider an extension to the linear cryptanalytic attack [4, 5] using multiple linear approximations. This offers a slight improvement in the efficiency of an attack on DES but more importantly, it is generally applicable and in certain circumstances it might well be extremely effective in reducing the amount of data required by a cryptanalyst for a successful attack on a block cipher using linear cryptanalysis.

Our paper is organized as follows. We briefly describe the technique of linear cryptanalysis, and then we present an adaptation of these methods which allows us to use multiple linear approximations. After providing theoretical estimates for the performance of our techniques we present experimental evidence that supports our claims. We then consider some of the implications of our work and draw our conclusions.

Y.G. Desmedt (Ed.): Advances in Cryptology - CRYPTO '94, LNCS 839, pp. 26-39, 1994.

2 Linear Cryptanalysis

We shall assume familiarity with the original paper by Matsui [4] which introduced the linear cryptanalytic attack on DES; we shall also refer to the more recent paper due to Matsui [5].

The basic idea behind linear cryptanalysis is to find some linear approximation to the action of the iterated block cipher which connects together, in one expression, some bits of the plaintext $P_{i_1} \ldots P_{i_a}$, ciphertext $C_{j_1} \ldots C_{j_b}$ and key $K_{k_1} \ldots K_{k_u}$. We shall write $P_{i_1} \oplus \ldots \oplus P_{i_a}$ as $P[\chi_P]$ and we can write a single linear approximation as

$$P[\chi_P] \oplus C[\chi_C] = K[\chi_K]. \tag{1}$$

If equation 1 is correct with probability $p = \frac{1}{2} + \epsilon$ for randomly chosen plaintext and a fixed key, then we say that it has *bias* ϵ. By collecting *known* plaintext/ciphertext pairs the cryptanalyst can make a guess for the value of $K[\chi_K]$; provided $\epsilon \neq 0$ the guess becomes more reliable as the cryptanalyst collects more plaintext/ciphertext pairs.

We note that the expected value of the left side will be $\frac{1}{2} + \epsilon$ if $K[\chi_K] = 1$ and $\frac{1}{2} - \epsilon$ if $K[\chi_K] = 0$. Under certain heuristic assumptions the cryptanalyst is attempting to distinguish a distribution with mean $\frac{1}{2} + \epsilon$ and variance $\frac{1}{4} - \epsilon^2$ from one with mean $\frac{1}{2} - \epsilon$ and variance $\frac{1}{4} - \epsilon^2$ [2]. The cryptanalyst has to take sufficiently many plaintext/ciphertext pairs to be confident that this distinction is being correctly made and so the smaller the value of ϵ, the more plaintext/ciphertext pairs are required to give the same level of confidence that the identification is correct.

The basic algorithm that allows the cryptanalyst to deduce one bit of key information from a single linear approximation is *Algorithm 1* in [4, 5].

Algorithm 1

Suppose that equation 1, $P[\chi_P] \oplus C[\chi_C] = K[\chi_K]$, is correct with probability $p = \frac{1}{2} + \epsilon$.

Step 1 Let T be the number of plaintext/ciphertext pairs such that the left side of equation 1 is equal to 0 and let N be the total number of pairs.

Step 2 If $T > N/2$

- then guess $K[\chi_K] = 0$ (when $\epsilon > 0$) or 1 (when $\epsilon < 0$),
- else guess $K[\chi_K] = 1$ (when $\epsilon > 0$) or 0 (when $\epsilon < 0$).

Of more practical importance is *Algorithm 2* [5] which allows a cryptanalyst attacking DES to recover up to 13 bits of key information in the full 16-round DES. More generally, for an r-round Feistel cipher we approximate $(r-2)$ iterations of the F-function from the second to the $(r-1)^{\text{th}}$ round using some linear approximation while guessing the subkey bits that are relevant to the first and final rounds. To keep the number of candidates small, it is advisable to consider linear approximations that involve guessing few subkey bits in the first and r^{th}

round. In the case of DES, this is conveniently achieved by guessing the subkeys relevant to a single S-box in the outer rounds.

Following Matsui we write the parity of the relevant output bits from the F-function in round one, which is dependent on subkey K_1 and plaintext block P_L, as $F_1(P_L, K_1)[\chi_{F_1}]$; we use similar notation for the output of the F-function in round r. The relevant linear approximation can be written as follows:

$$P[\chi_P] \oplus C[\chi_C] \oplus F_1(P_L, K_1)[\chi_{F_1}] \oplus F_r(C_L, K_r)[\chi_{F_r}] = K[\chi_K]. \qquad (2)$$

Suppose that equation 2 is correct with probability $p = \frac{1}{2} + \epsilon$.

Algorithm 2

Step 1 Let $K_1^{(g)}$ $(g = 1, 2, \ldots)$ and $K_r^{(h)}$ $(h = 1, 2, \ldots)$ be possible candidates for K_1 and K_r respectively. Then for each pair $(K_1^{(g)}, K_r^{(h)})$, let $T_{g,h}$ be the number of plaintexts such that the left side of equation 2 is equal to 0 when K_1 is replaced by $K_1^{(g)}$ and K_r by $K_r^{(h)}$. Let N be the total number of plaintexts.

Step 2 Let T_{max} be the maximum value and T_{min} be the minimum value of all $T_{g,h}$'s.

- If $|T_{max} - N/2| > |T_{min} - N/2|$, adopt the key candidate corresponding to T_{max} and guess $K[\chi_K] = 0$ (when $\epsilon > 0$) or 1 (when $\epsilon < 0$).
- If $|T_{max} - N/2| < |T_{min} - N/2|$, adopt the key candidate corresponding to T_{min} and guess $K[\chi_K] = 1$ (when $\epsilon > 0$) or 0 (when $\epsilon < 0$).

Algorithm 2 provides a guess for the relevant subkey bits of K_1 and K_r together with the additional bit of subkey information obtained from the linear approximation. Matsui [5] gives both experimental and theoretical justification for the belief that by taking $8\epsilon^{-2}$ known plaintext/ciphertext pairs *Algorithm 2* is correct with high probability.

To obtain more subkey bits, another linear approximation can be used or, for a Feistel cipher, the roles of the ciphertext and plaintext can be reversed and *Algorithm 2* applied again.

In the particular case of DES, Matsui [5] uses the second approach to obtain 26 bits of subkey information in total. Fortuitously, the key schedule in DES ensures that these 26 bits of subkey information correspond to 26 distinct bits of key information. The remaining unknown 30 bits of key can be derived using an exhaustive search.

3 Multiple Linear Approximations

There are many different linear approximations to a block cipher over a given number of rounds. Suppose we have n linear approximations which involve the same key bits but differ in the plaintext and ciphertext bits that they use.

Matsui's *Algorithm 1* can be used on any of the n individual linear approximations to define a statistic T_i, $1 \leq i \leq n$, with a certain bias and variance;

the success of *Algorithm 1* depends on both. However, we now show that it is possible to use more than one linear approximation at the same time. In short, we work harder on the information we already have rather than calling for more plaintext/ciphertext pairs.

Suppose the i^{th} linear approximation for $1 \leq i \leq n$ has the following form:

$$P[\chi_P^i] \oplus C[\chi_C^i] = K[\chi_K]. \tag{3}$$

For the sake of analysis, we will suppose that each bias ϵ_i is positive; changes can easily be made to the data collecting phase to ensure that this condition holds.

Algorithm 1M

Step 1 For $1 \leq i \leq n$ let T_i be the number of plaintext/ciphertext pairs such that the left side of equation 3 is equal to 0. Let N denote the total number of plaintexts.

Step 2 For some set of weights $a_1, \ldots a_n$ where $\sum_{i=1}^{n} a_i = 1$ calculate

$$U = \sum_{i=1}^{n} a_i T_i.$$

Step 3 If $U > \frac{N}{2}$ then guess $K[\chi_K] = 0$, else guess $K[\chi_K] = 1$.

The analysis below will show that we have introduced a new statistic U (defined in terms of the T_i, $1 \leq i \leq n$) which has a bias comparable to the T_i's but a reduced variance. Because of this reduced variance the attack requires fewer known plaintexts. We will also see in Lemma 2 that the success rate of *Algorithm 1M* is optimized if the weights are defined as $a_i = \epsilon_i / \sum \epsilon_i$.

3.1 Analysis of *Algorithm 1M*

For each linear approximation define x_i to be a random variable whose value is 0 when the left side of the i^{th} linear approximation is equal to 1 and 1 when the left side is equal to 0. By making this definition we see that sum of x_i for the i^{th} linear approximation over all plaintext/ciphertext pairs in the experiment is equal to T_i.

We observe that for a good block cipher the value of the left sides of two different linear approximations will be essentially independent since the two left sides differ by a combination of plaintext and ciphertext bits. This observation is supported by the experimental work presented in Section 4, and leads to the following assumption:

Assumption 1 *For all i and j with $i \neq j$, $x_i = x_j$ with probability $\frac{1}{2}$, where the probability is taken over randomly chosen plaintexts.*

We now establish the values for the mean and variance of the statistic U.

Lemma 1. *Let* $y = \sum_{i=1}^{n} a_i x_i$ *with* $\sum_{i=1}^{n} a_i = 1$. *Then under Assumption 1 we have* $E[y] = \frac{1}{2} + \sum_{i=1}^{n} a_i \epsilon_i$ *if* $K[\chi_K] = 0$ *and* $E[y] = \frac{1}{2} - \sum_{i=1}^{n} a_i \epsilon_i$ *if* $K[\chi_K] = 1$. *The variance satisfies*

$$\sigma_y^2 = \sum_{i=1}^{n} \frac{a_i^2}{4} - \left(\sum_{i=1}^{n} a_i \epsilon_i \right)^2 .$$

Proof. See Appendix A. □

Following Matsui and noting that each value of the statistic $U = \sum a_i T_i$ can also be expressed as a sum of independent values $y = \sum a_i x_i$, we will make the following assumption which is consistent with the experimental results presented in Section 4.

Assumption 2 *The distribution of the statistic* $U = \sum_{i=1}^{n} a_i T_i$ *can be accurately modeled using a normal distribution.*

Our goal is to maximize the distance between $N/2$ and $E[U] = NE[y]$ in terms of the standard deviation $\sigma_U = \sqrt{N}\sigma_y$. This increases our success rate for a fixed N, or alternatively, allows us to use a smaller N while maintaining the same level of success.

Lemma 2. *Under Assumptions 1 and 2, with the statistic* U *defined as in Algorithm 1M, the distance* $|N/2 - E[U]|/\sigma_U$ *is maximized for a given* N *when the weights* a_i *are proportional to the biases* ϵ_i.

Proof. See Appendix A. □

The maximum distance is easily calculated as $2\sqrt{N}\sqrt{(\sum \epsilon_i^2)/(1 - 4\sum \epsilon_i^2)}$. Adopting the conventional notation of $\Phi(\cdot)$ for the normal cumulative distribution function [2], this leads to the following theorem:

Theorem 3. *Under Assumptions 1 and 2, the success rate of Algorithm 1M, with optimal weights* a_i, *is*

$$\Phi\left(2\sqrt{N} \sqrt{\frac{\sum_{i=1}^{n} \epsilon_i^2}{1 - 4\sum_{i=1}^{n} \epsilon_i^2}} \right) .$$

Proof. Follows directly. □

When $\sum \epsilon_i^2$ is small, we approximate the success rate as $\Phi(2\sqrt{N}\sqrt{\sum \epsilon_i^2})$, a generalization of Matsui's single-approximation success rate of $\Phi(2\sqrt{N}\epsilon)$.

As an illustration of the improvement possible, suppose that we have n linear approximations all involving the same key bits and all having the same bias ϵ. Then using *Algorithm 1M* and N' known plaintext/ciphertext pairs the success rate is given by $\Phi\left(2\sqrt{N'}\sqrt{\sum \epsilon_i^2}\right) = \Phi\left(2\sqrt{N'}\sqrt{n}\epsilon\right)$. Using *Algorithm 1* with

a single linear approximation, the success rate is $\Phi\left(2\sqrt{N}\epsilon\right)$. We see that if $N' = N/n$, then there is an n-fold reduction in the amount of data and yet the two algorithms have the same success rate.

It is somewhat artificial to assume that there are n linear approximations involving the same key bits, all correct with the same probability. Later we will outline some extensions which allow us to relax these conditions on the use of multiple linear approximations. However, we can summarize our results so far by saying that the use of several carefully chosen linear approximations concurrently will lead to a reduction in the amount of data required for a linear cryptanalytic attack.

3.2 *Algorithm 2M*

We now introduce an extension to *Algorithm 2* [4, 5] which uses multiple linear approximations. For an r-round Feistel cipher we approximate $(r-2)$ iterations of the F-function from the second to the $(r-1)^{\text{th}}$ round using n linear approximations while we still make guesses for the subkey bits needed to extend through the first and final rounds. Note that to keep the number of candidates small, the approximations should all involve the same guessed subkey bits in round one, as well as in round r. Following the notation established for *Algorithm 2* we can write the i^{th} linear approximation as follows:

$$P[\chi_P^i] \oplus C[\chi_C^i] \oplus F_1(P_L, K_1)[\chi_{F_1}^i] \oplus F_r(C_L, K_r)[\chi_{F_r}^i] = K[\chi_K]. \qquad (4)$$

We will again suppose, without loss of generality, that each bias ϵ_i is positive.

Step 1 Let $K_1^{(g)}$ $(g = 1, 2, \ldots)$ and $K_r^{(h)}$ $(h = 1, 2, \ldots)$ be possible candidates for K_1 and K_r respectively. Then for each pair $(K_1^{(g)}, K_r^{(h)})$ and each linear approximation i let $T_{g,h}^i$ be the number of plaintexts such that the left side of equation 4 is equal to 0 when K_1 is replaced by $K_1^{(g)}$ and K_r by $K_r^{(h)}$. Let N be the total number of plaintexts.

Step 2 Let $a_i = \epsilon_i / \sum_{i=1}^n \epsilon_i$. Calculate, for each g, h,

$$U_{g,h} = \sum_{i=1}^{n} a_i T_{g,h}^i$$

Step 3 Let U_{max} be the maximum value and U_{min} be the minimum value of all $U_{g,h}$'s.

- If $|U_{max} - N/2| > |U_{min} - N/2|$, adopt the key candidate corresponding to U_{max} and guess $K[\chi_K] = 0$.
- If $|U_{max} - N/2| < |U_{min} - N/2|$, adopt the key candidate corresponding to U_{min} and guess $K[\chi_K] = 1$.

Algorithm 2M is a generalization of *Algorithm 1M* and we expect it to be successful for essentially the same reasons. The main issue is that we are replacing the statistic defined using *Algorithm 2* with one that has a smaller variance. We

know that the variance is reduced for the correct key guess; the complicating issue is what happens to the statistics $U_{g,h}$ for incorrect key guesses. By Lemma 1 the use of multiple linear approximations ensures that the variance of any statistic, even those for incorrect key guesses, will be reduced. This gives us one reason to expect a higher success rate with *Algorithm 2M* for a given number of plaintext/ciphertext pairs.

Another reason is due to a cancellation property. Let $P^i_{g,h}$ be the probability for a random plaintext/ciphertext pair that an incorrect key guess gives the same answer as a correct key guess; that is, the probability that

$$F_1(P_L, K_1^{(g)})[\chi^i_{F_1}] \oplus F_r(P_L, K_r^{(h)})[\chi^i_{F_r}] = F_1(P_L, K_1)[\chi^i_{F_1}] \oplus F_r(P_L, K_r)[\chi^i_{F_r}].$$

Then it is easy to show that

$$E[U_{g,h}] = \frac{N}{2} + N \sum_{i=1}^{n} a_i \epsilon_i \delta^i_{g,h}$$

where $\delta^i_{g,h}$ is a "correlation coefficient" defined as $\delta^i_{g,h} = 2P^i_{g,h} - 1$. This coefficient takes values between -1 and $+1$.

The greater the magnitude of the difference between $E[U_{g,h}]$ and $N/2$, the greater the likelihood that the key guess $(K_1^{(g)}, K_r^{(h)})$ is adopted. However, since a correlation coefficient $\delta^i_{g,h}$ can take both positive and negative values for an incorrect key guess, but is always 1 for the correct key guess, the correct guess is at a significant advantage. We will explore this phenomenon more closely in subsequent work.

Part of Section 4 contains results on the performance of *Algorithm 2M*.

4 Experimental Results

In this section we report the findings from several experiments performed on small-round versions of DES. We chose small-round versions because we wish to present confirmation of the performance of our algorithms and we feel that these examples are particularly illustrative.

4.1 Confirming the assumptions and *Algorithm 1M*

The aim of the first set of experiments was to substantiate the assumptions made during Section 3.1 and to test the theoretical predictions made about the use of multiple linear approximations. Although the experiments were conducted on a small scale, they provide good corroborating evidence for the validity of both the assumptions and the theoretical results. It is reasonable to assume that these results would extend to a larger scale.

We used a five-round version of DES with two linear approximations identified using the notation in [5] as $-$`ACD`$-$ and $-$`DCA`$-$. Each linear approximation

holds with bias $\epsilon = 25 \times 2^{-12} \approx 6.104 \times 10^{-3}$. Let K_i denote the 48-bit subkey used at round i, P_L the low 32 bits of plaintext and C_L the low 32 bits of ciphertext. Then we have

$$P_1 : P_L[7, 18, 24, 29] \oplus C_L[7, 18, 24] = K_2[22] \oplus K_3[44] \oplus K_4[22]$$
$$P_2 : P_L[7, 18, 24] \oplus C_L[7, 18, 24, 29] = K_2[22] \oplus K_3[44] \oplus K_4[22]$$

First we consider Assumption 1. Using $2,000,000$ plaintext/ciphertext pairs and linear approximations P_1 and P_2 we found that the outcome of the two relations agreed $999,351$ times and disagreed the remaining $1,000,649$ times; this gives us some confidence that Assumption 1 is reasonable.

We next completed 100 attempts at the linear cryptanalysis of five-round DES using *Algorithm 1M* on an increasing number of plaintext/ciphertext pairs; these numbers were chosen so that the number of plaintext/ciphertext pairs was roughly $\frac{1}{8}\epsilon^{-2}, \frac{1}{4}\epsilon^{-2}, \frac{1}{2}\epsilon^{-2}$ and ϵ^{-2}. The success rates achieved in practice using P_1 and P_2 individually and then jointly are presented below, together with the results predicted by the theory for *Algorithm 1* using individual linear approximations and for *Algorithm 1M* using multiple linear approximations.

	experimental results				*theoretical predictions*			
number of pairs	3,356	6,711	13,422	26,844	3,356	6,711	13,422	26,844
P_1	81%	86%	94%	99%	76%	84%	92%	98%
P_2	75%	88%	92%	99%	76%	84%	92%	98%
using P_1 *and* P_2	92%	95%	98%	100%	84%	92%	98%	100%

These results confirm Theorem 3: it is clear that the use of two linear approximations gives success rates comparable to individual linear approximations with half as many plaintext/ciphertext pairs.

4.2 Confirming *Algorithm 2M*

The aim of the second set of experiments was to confirm the expected behavior of *Algorithm 2M*. The experiments consisted of attacking a seven-round version of DES by using the five-round linear approximations P_1 and P_2 from the previous experiment for rounds two through six, and guessing the 12 subkey bits used in the first and seventh rounds. The results are presented below. We note that the same success rate achieved using either of the single linear approximations on their own could be achieved by using both linear approximations together with half as many plaintext/ciphertext pairs.

	experimental results			
number of pairs	13,422	26,844	53,688	107,376
P_1	3%	2%	17%	51%
P_2	1%	1%	15%	45%
using P_1 *and* P_2	4%	13%	46%	94%

5 Effectiveness and Extensions

In an attack on DES it is perhaps surprising how little advantage is gained at present using multiple linear approximations. An exhaustive search reveals that there are 10,006 14-round linear approximations involving a single S-box at each round with bias $|\epsilon_i| > 10^{-8}$. We found that $\sum_{i=1}^{10,006} \epsilon_i^2 \approx 1.23 \times 10^{-11}$. With our present techniques we can use only a small fraction of these approximations, giving only a minor improvement in the number of plaintext/ciphertext pairs required.

Even if we could use all these linear approximations at the same time, it would only result in a reduction by a factor of about 38 in the number of plaintext/ciphertext pairs, since for the "best" linear approximation, $\epsilon^2 \approx 3.22 \times 10^{-13}$. Future research will undoubtedly reveal whether techniques exist using multiple linear approximations which can begin to deliver such improvements.

Note that there may well be different linear approximations involving different subkey bits which, because of the key schedule used in the block cipher, correspond to the same key-bits in the user-provided key. In such a case these linear approximations would each provide a guess for the same bit of key information, and could be used together with *Algorithm 1M* or *2M*. Unfortunately the existence of such linear approximations depends closely on the key schedule in the cipher.

Our aim is to remove the assumption that all the linear approximations use the same subkey bits. Our approach is to modify the techniques we have, without regard for the key schedule, to use good linear approximations which potentially involve different subkey bits. At this point we present only an extension to *Algorithm 1M*.

Suppose we have n linear approximations and the i^{th} linear approximation has the form

$$P[\chi_P^i] \oplus C[\chi_C^i] = K[\chi_K^i]. \tag{5}$$

To consider the approximations together, we must first guess, for each j, $2 \leq j \leq n$, whether $K[\chi_K^j]$ and $K[\chi_K^i]$ are equal or not. There are at most 2^{n-1} guesses and for each guess we obtain n linear approximations of the form

$$P[\chi_P^i] \oplus C[\chi_C^i] \oplus \Delta^i = K[\chi_K^1] \tag{6}$$

where $\Delta^1 = 0$ and Δ^i for $i > 1$ depends on the guess.

At this point, for each guess, we can determine by *Algorithm 1M* whether $K[\chi_K^1] = 0$ or 1, where the statistic T_i in *Algorithm 1M* is taken from the left side of equation 6. Note that in practice one would not repeat *Algorithm 1M* for each guess; instead one would consider up to 2^{n-1} ways of combining the statistics T_i for $1 \leq i \leq n$.

This approach does not determine which guess is correct, but for each guess it does give an estimate for the value of $K[\chi_K^1]$. In effect this halves the exhaustive search space. We need only have a high success rate for the value of $K[\chi_K^1]$ for the correct guess. Analysis similar to that for *Algorithm 1M* suggests that fewer

plaintext/ciphertext pairs are required for a given success rate than if we only used a single approximation.

As an example of this approach, consider the following linear approximations to six-round DES. Following Matsui's notation, let P_1 represent the linear approximation A-ACD- while P_2 represents E-DCA-. These hold with bias 3.81×10^{-3} and 3.05×10^{-3} respectively.

$$P_1 : P_H[7,18,24,29] \oplus P_L[15] \oplus C_L[7,18,24] = K_1[22] \oplus K_3[22] \oplus K_4[44] \oplus K_5[22] \quad (7)$$

$$P_2 : P_H[7,18,24] \oplus P_L[12,16] \oplus C_L[7,18,24,29] = K_1[19,23] \oplus K_3[22] \oplus K_4[44] \oplus K_5[22] \quad (8)$$

We can use both linear approximations in *Algorithm 1M* while guessing whether the right sides are equal or not. The results of this experiment are provided below. The amount of data in these experiments was chosen so that the number of plaintext/ciphertext pairs was roughly $\frac{1}{8}\epsilon^{-2}$, $\frac{1}{4}\epsilon^{-2}$, $\frac{1}{2}\epsilon^{-2}$ and ϵ^{-2} where $\epsilon = 3.81 \times 10^{-3}$. The success rates provided in the table below are for the correct guess, which in our experiment was that the right sides are equal.

More analysis of this technique will be presented in future work.

	experimental results				*theoretical predictions*			
number of pairs	8,590	17,180	34,360	68,720	8,590	17,180	34,360	68,720
P_1	81%	82%	92%	100%	76%	84%	92%	98%
P_2	73%	84%	86%	96%	71%	79%	87%	95%
using P_1 *and* P_2	89%	93%	99%	100%	82%	90%	96%	99%

6 Implications

If we consider DES-like ciphers generally, we can imagine situations where the use of multiple linear approximations might well be very significant. If an S-box has b output bits, then there are potentially $2^b - 1$ useful linear approximations using the same input mask, and hence the same subkey bits. We note that there are no useful variants of the trivial one-round linear approximation (denoted as – by Matsui).

The Feistel structure of a DES-like cipher means that we can consider $2^b - 1$ variations on the left half of the plaintext and $2^b - 1$ variations on the right half which would provide alternative linear approximations on exactly the same subkey bits. Since the trivial one-round linear approximation has no useful variations, the best linear approximations (which we would expect to use the trivial one-round linear approximation) might not allow consideration of all these variations.

One of the most interesting cases is a DES-like block cipher with larger S-boxes. Our results show that the extra linear approximations admitted by the use of larger S-boxes might increase the vulnerability of the block cipher to some form of linear cryptanalysis using multiple linear approximations. More research

is required to assess the significance of this development but we note that it should be an important consideration; after the work of O'Conner [8] larger S-boxes have been proposed as a way of increasing the resistance of a block cipher to differential cryptanalysis [3].

Another consequence of our work is that in assessing the resistance of a block cipher to linear cryptanalysis it is not sufficient to consider the best linear approximation in isolation; one must also take account of other linear approximations which might be used at the same time. We suggest that some of the preliminary work in assessing the practical security of a cipher against linear cryptanalytic attack [3] should be broadened since these estimates of strength are made under the assumption that the best attack is completed using a single linear approximation. Under this assumption a lower bound on the complexity of a linear cryptanalytic attack on a Feistel cipher is obtained by considering (i) the bias of the best non-trivial one-round linear approximation and (ii) the number of rounds in the cipher.

By considering a theoretical best linear approximation one overlooks the possibility that several good linear approximations could be used concurrently to obtain a more efficient attack. We suggest that a more useful measure of practical security against linear cryptanalytic attack should also consider the use of multiple linear approximations.

7 Conclusions

We have presented an extension to the basic linear cryptanalytic attack which offers an improvement in the number of known plaintext/ciphertext pairs required for the linear cryptanalysis of a block cipher. While its effectiveness in an attack on DES is at present somewhat limited, it is a general technique which might have excellent results in the cryptanalysis of other less well designed block ciphers.

Importantly, we note that the use of larger S-boxes, which is sometimes recommended as a way of increasing the security of DES-like block ciphers, might in certain circumstances facilitate the use of linear cryptanalysis with multiple linear approximations. More research is needed to ascertain quite how significant a threat this might eventually be.

We believe that the use of multiple linear approximations is an important cryptanalytic tool and one which should be considered both in the design of block ciphers and in any attempt to provide a theoretical bound on the resistance of block ciphers to linear cryptanalysis.

References

1. E. Biham and A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard.* Springer-Verlag, New York, 1993.
2. A.W. Drake. *Fundamentals of Applied Probability Theory.* McGraw-Hill Book Company, New York, 1967.

3. L.R. Knudsen. Practically secure Feistel ciphers. In *Proceedings of Cambridge Security Workshop, December 1993*, Springer-Verlag, Berlin, To appear.
4. M. Matsui. Linear cryptanalysis method for DES cipher. In T. Helleseth, editor, *Advances in Cryptology — Eurocrypt '93*, pages 386–397, Springer-Verlag, Berlin, 1994.
5. M. Matsui. Linear cryptanalysis of DES cipher (I). January 1994. Preprint.
6. M. Matsui and A. Yamagishi. A new method for known plaintext attack of FEAL cipher. In R.A. Rueppel, editor, *Advances in Cryptology — Eurocrypt '92*, pages 81–91, Springer-Verlag, Berlin, 1992.
7. National Institute of Standards and Technology (NIST). *FIPS Publication 46-2: Data Encryption Standard.* December 30, 1993. Originally issued by National Bureau of Standards.
8. L. O'Conner. On the distribution of characteristics in bijective mappings. In T. Helleseth, editor, *Advances in Cryptology — Eurocrypt '93*, pages 360–370, Springer-Verlag, Berlin, 1994.
9. E. Okamoto. Personal communication. March 1994.
10. A. Shimizu and S. Miyaguchi. Fast data encipherment algorithm FEAL. In D. Chaum and W.L. Price, editors, *Advances in Cryptology — Eurocrypt '87*, pages 267–280, Springer-Verlag, Berlin, 1988.

A Proof of Lemmas 1 and 2

We need the following additional lemma for our proof of Lemma 1.

Lemma 4. *Under Assumption 1, $E[x_i x_j]$ is given by*

$$E[x_i x_j] = E[x_i]E[x_j] - \epsilon_i \epsilon_j .$$

Proof. Clearly $x_i x_j = 1$ if, and only if, $x_i = x_j = 1$. Thus $E[x_i x_j] = \frac{1}{2} E[x_i | x_i = x_j]$, since $x_i = x_j$ with probability $\frac{1}{2}$. If $K[\chi_K] = 0$, we know that

$$\frac{1}{2} E[x_i | x_i = x_j] + \frac{1}{2} E[x_i | x_i \neq x_j] = \frac{1}{2} + \epsilon_i$$

and

$$\frac{1}{2} E[x_j | x_i = x_j] + \frac{1}{2} E[x_j | x_i \neq x_j] = \frac{1}{2} + \epsilon_j .$$

It is not hard to solve for the individual expectations

$$E[x_i | x_i = x_j] = E[x_j | x_i = x_j] = \frac{1}{2} + \epsilon_i + \epsilon_j ,$$

so that

$$E[x_i x_j] = \frac{1}{4} + \frac{\epsilon_i + \epsilon_j}{2} = \left(\frac{1}{2} + \epsilon_i\right)\left(\frac{1}{2} + \epsilon_j\right) - \epsilon_i \epsilon_j$$

and the result follows. Analysis is similar if $K[\chi_K] = 1$. □

Proof of Lemma 1

Proof. It is straightforward to verify that $E[y] = \frac{1}{2} \pm \sum a_i \epsilon_i$ depending on $K[\chi_K]$. The calculation of the variance σ_y^2 is more involved.

$$\begin{aligned}
\sigma_y^2 &= E[y^2] - E[y]^2 \\
&= E\left[\left(\sum_{i=1}^n a_i x_i\right)^2\right] - \left(E[y]\right)^2 \\
&= E\left[\sum_{i=1}^n a_i^2 x_i^2 + 2\sum_{i=1}^n \sum_{j=i+1}^n a_i a_j x_i x_j\right] - \left(\sum_{i=1}^n a_i E[x_i]\right)^2 \\
&= \sum_{i=1}^n a_i^2 E\left[x_i^2\right] + 2\sum_{i=1}^n \sum_{j=i+1}^n a_i a_j E[x_i x_j] \\
&\quad - \sum_{i=1}^n a_i^2 \left(E[x_i]\right)^2 - 2\sum_{i=1}^n \sum_{j=i+1}^n a_i a_j E[x_i] E[x_j] \\
&= \sum_{i=1}^n a_i^2 \left(E[x_i^2] - \left(E[x_i]\right)^2\right) + 2\sum_{i=1}^n \sum_{j=i+1}^n a_i a_j \left(E[x_i x_j] - E[x_i] E[x_j]\right) \\
&= \sum_{i=1}^n \frac{a_i^2}{4} - \sum_{i=1}^n a_i^2 \epsilon_i^2 + 2\sum_{i=1}^n \sum_{j=i+1}^n a_i a_j \left(E[x_i x_j] - E[x_i] E[x_j]\right).
\end{aligned}$$

Under Assumption 1 we established in Lemma 4 that

$$E[x_i x_j] = E[x_i] E[x_j] - \epsilon_i \epsilon_j .$$

This then gives us

$$\begin{aligned}
\sigma_y^2 &= \sum_{i=1}^n \frac{a_i^2}{4} - \sum_{i=1}^n a_i^2 \epsilon_i^2 - 2\sum_{i=1}^n \sum_{j=i+1}^n a_i a_j \epsilon_i \epsilon_j \\
&= \sum_{i=1}^n \frac{a_i^2}{4} - \left(\sum_{i=1}^n a_i \epsilon_i\right)^2 .
\end{aligned}$$

□

Proof of Lemma 2

Proof. The distance $|N/2 - E[U]|/\sigma_U$ is maximized when $(N/2 - E[U])^2/\sigma_U^2$ is maximized. Expanding $E[U]$ and σ_U^2, we wish to maximize

$$\begin{aligned}
\frac{(N/2 - N E[y])^2}{N \sigma_y^2} &= N \frac{\left(\sum_{i=1}^n a_i \epsilon_i\right)^2}{(1/4)\sum_{i=1}^n a_i^2 - \left(\sum_{i=1}^n a_i \epsilon_i\right)^2} \\
&= N \frac{4}{\left(\sum_{i=1}^n a_i^2\right)/\left(\sum_{i=1}^n a_i \epsilon_i\right)^2 - 4}.
\end{aligned}$$

It is easy to show that this is maximized when $(\sum a_i\epsilon_i)^2/(\sum a_i^2)$ is maximized; note that the denominator must be positive since σ_y^2 must be. We now give an upper bound of $\sum \epsilon_i^2$ on this fraction, by contradiction.

Suppose that the fraction exceeds the upper bound. Then we must have

$$(\sum_{i=1}^{n} a_i\epsilon_i)^2 > (\sum_{i=1}^{n} a_i^2)(\sum_{i=1}^{n} \epsilon_i^2).$$

Expanding the terms, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\epsilon_i a_j\epsilon_j > \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^2\epsilon_j^2.$$

Thus we must have, for some i and j,

$$2a_i\epsilon_i a_j\epsilon_j > a_i^2\epsilon_j^2 + a_j^2\epsilon_i^2,$$

or $(a_i\epsilon_j - a_j\epsilon_i)^2 < 0$, which is a contradiction.

The upper bound is achieved when $a_i\epsilon_j - a_j\epsilon_i = 0$ for all i and j, i.e., when a_i/ϵ_i is constant. □