

Overestimates for the Gain of Multiple Linear Approximations

S. Murphy

Information Security Group,
Department of Mathematics,
Royal Holloway, University of London,
Egham, Surrey TW20 0EX, U.K.

Abstract. We show that Corollary 1 of “On Multiple Linear Approximations” (Crypto 2004 – LNCS 3152) is incorrect. In particular, the value given for the gain by Corollary 1 is likely to be a significant overestimate of this quantity. Thus any data requirements for linear cryptanalysis with multiple linear approximations based on this value for the gain are highly questionable.

Keywords. Linear cryptanalysis, multiple linear approximations, gain, Jensen’s inequality.

1 Introduction

Linear cryptanalysis [4] of a block cipher in its basic form uses a linear approximation of the form

$$\alpha^T \begin{pmatrix} \mathbf{p} \\ \mathbf{c} \end{pmatrix} = k \text{ with probability } \frac{1}{2}(1 + \epsilon),$$

where α is a data mask, k is one bit of key information, \mathbf{p} is a plaintext and \mathbf{c} is a corresponding ciphertext. The value ϵ is known as the *imbalance* or *correlation* (twice the *bias*) of the linear approximation. If $\epsilon \neq 0$, then it is possible to estimate the key bit k reasonably accurately if the number N of plaintext-ciphertext pairs is at least ϵ^{-2} [4].

Enhanced forms of linear cryptanalysis [2, 3] use a collection of m such linear approximations. Such a situation with multiple linear approximations is also considered by [1], where the *gain* of such a linear cryptanalysis is defined. The gain is an attempt to quantify the advantage of a such a linear cryptanalysis over exhaustive search.

This paper is concerned with the values given for the gain by [1]. In particular, we show that the value for the gain given there by *Corollary 1* generally greatly exceeds the value for the gain given there by *Theorem 1*.

2 Multiple Linear Approximations

We consider a linear cryptanalysis based on N plaintext-ciphertext pairs. We suppose that we have m linear approximations

$$\alpha_i^T \begin{pmatrix} \mathbf{P} \\ \mathbf{c} \end{pmatrix} = k_i \text{ with probability } \frac{1}{2}(1 + \epsilon_i)$$

for distinct data masks α_i , individual bits of key information k_i and imbalances ϵ_i ($i = 1, \dots, m$). The *capacity* \bar{c}^2 of this collection of linear approximations is given by Definition 2 of [1] to be $\bar{c}^2 = \sum_{i=1}^m \epsilon_i^2$.

For simplicity, we suppose that the m key bits k_1, \dots, k_m give m bits of information about the block cipher key. We let $\mathbf{z} = (k_1, \dots, k_m)^T$ denote the *key class*, and we denote the set of all key classes by \mathcal{Z} , so $\mathcal{Z} = \mathbb{Z}_2^m$ and $|\mathcal{Z}| = 2^m$. We let \mathbf{z}^* denote the key class containing the true key, and, without loss of generality, we suppose that $\mathbf{z}^* = 0$. We let $\mathcal{Z}^* = \mathcal{Z} \setminus \{\mathbf{z}^*\} = \mathbb{Z}_2^m \setminus \{0\}$ denote the set of key classes not containing the true key, so $|\mathcal{Z}^*| = 2^m - 1$. We denote the m -dimensional *imbalance vector* corresponding to key class \mathbf{z} by \mathbf{c}_z , so

$$\mathbf{c}_z = ((-1)^{z_1} \epsilon_1, \dots, (-1)^{z_m} \epsilon_m)^T.$$

We note that the squared distance from such an imbalance vector to the imbalance vector for the true key class is given by

$$|\mathbf{c}_z - \mathbf{c}_{z^*}|^2 = |\mathbf{c}_z - \mathbf{c}_0|^2 = \left| -2(z_1 \epsilon_1, \dots, z_m \epsilon_m)^T \right|^2 = 4 \sum_{i=1}^m z_i^2 \epsilon_i^2.$$

3 Mathematical Concepts used to Define Gain Values

The values given for the gain by [1] can be expressed in terms of two functions, g and H_m , and a random variable X , which we now define.

The function g on the positive real numbers is defined by

$$g(x) = \phi \left(-\frac{1}{2} N^{\frac{1}{2}} x^{\frac{1}{2}} \right),$$

where ϕ denotes the cumulative distribution function for a standard normal $N(0; 1)$ random variable. We note that $g(x)$ is a convex function of x for $x > 0$ as

$$g''(x) = \frac{1}{32} \frac{1}{\sqrt{2\pi}} N^{\frac{1}{2}} e^{-\frac{Nx}{8}} (N + 4x^{-1}) x^{-\frac{1}{2}} > 0 \text{ for } x > 0.$$

The function H_m on the positive real numbers is defined by

$$H_m(x) = -\log_2 (2(1 - 2^{-m})x + 2^{-m}) = -\log_2 \left[2 \frac{|Z^*|}{|Z|} x + \frac{1}{|Z|} \right].$$

We note that $H_m(x)$ is a decreasing function of x for $x > 0$ as

$$H'_m(x) = -\frac{1}{\log 2} \left(\frac{2(1 - 2^{-m})}{2(1 - 2^{-m})x + 2^{-m}} \right) < 0 \text{ for } x > 0.$$

The random variable X is defined by

$$X = |\mathbf{c}_{\mathbf{z}} - \mathbf{c}_{\mathbf{z}^*}|^2 \text{ with probability } |Z^*|^{-1} = (2^m - 1)^{-1} \text{ for } \mathbf{z} \in Z^*.$$

Thus X is the random variable giving the squared distance of an imbalance vector for an incorrect key class from the imbalance vector for the true key class.

4 Comparison of Values for the Gain

We now compare the two values given for the gain in *Theorem 1* and *Corollary 1* of [1]. We show in Appendix A that the value γ for the gain given by *Theorem 1* is given by

$$\gamma = H_m(\mathbf{E}[g(X)]).$$

We show in Appendix B that the value $\tilde{\gamma}$ for the gain given by *Corollary 1* is given by

$$\tilde{\gamma} = H_m(g(2\bar{c}^2)) = H_m(g((1 - 2^{-m})\mathbf{E}[X])).$$

However, this value $\tilde{\gamma}$ for the gain can be well approximated by $\hat{\gamma}$, where

$$\hat{\gamma} = H_m(g(\mathbf{E}[X])).$$

We now use Jensen's inequality [6] to compare γ and $\hat{\gamma}$. As g is a convex function of the positive real numbers, Jensen's inequality shows that

$$g(\mathbf{E}[X]) \leq \mathbf{E}[g(X)].$$

Furthermore H_m is a decreasing function of the positive real numbers, so

$$\hat{\gamma} = H_m(g(\mathbf{E}[X])) \geq H_m(\mathbf{E}[g(X)]) = \gamma.$$

However, $\tilde{\gamma}$ is usually extremely well-approximated by $\hat{\gamma}$, so giving Lemma 1.

Lemma 1. The value $\tilde{\gamma}$ for the gain given by *Corollary 1* generally exceeds the value γ given for the gain by *Theorem 1*.

5 Example Values for the Gain

The important issue in the use of *Corollary 1* of [1] to give the gain is whether the overestimate of γ by $\tilde{\gamma}$ referred to in Lemma 1 gives a significant error in the value of the gain. We show by giving an example that it is indeed generally the case that the use of $\tilde{\gamma}$ given in *Corollary 1* gives a large overestimate of the gain γ given by *Theorem 1*.

For simplicity, we assume that all m linear approximations have the same imbalance ϵ , that is $\epsilon_1 = \dots = \epsilon_m = \epsilon$. The capacity of such a collection of linear approximations is clearly $\bar{c}^2 = m\epsilon^2$. In this situation, using the result given in Section 2, we have

$$|\mathbf{c}_{\mathbf{z}} - \mathbf{c}_{\mathbf{z}^*}|^2 = 4\epsilon^2 \sum_{i=1}^m z_i^2 = 4\epsilon^2 |\mathbf{z}|^2.$$

As there are $\binom{m}{l}$ such vectors $\mathbf{z} \in \mathbb{Z}_2^m$ with $|\mathbf{z}|^2 = l$, the random variable X is given by

$$X = 4\epsilon^2 l \text{ with probability } \binom{m}{l} (2^m - 1)^{-1} \quad [l = 1, \dots, m].$$

Thus X is a multiple of a censored $\text{Bin}(m, \frac{1}{2})$ random variable with 0 removed, so the mean of X is given by $\mathbf{E}[X] = 4\epsilon^2 \frac{m}{2} \frac{2^m}{2^m - 1}$. We therefore obtain $g(\mathbf{E}[X])$, used to define $\hat{\gamma}$, as

$$\begin{aligned} g(\mathbf{E}[X]) &= \phi \left(-\frac{1}{2} N^{\frac{1}{2}} \epsilon \left(2m \frac{2^m}{2^m - 1} \right)^{\frac{1}{2}} \right) \\ &= \phi \left(- (N\epsilon^2)^{\frac{1}{2}} \left(\frac{m}{2} \right)^{\frac{1}{2}} \left(\frac{2^m}{2^m - 1} \right)^{\frac{1}{2}} \right). \end{aligned}$$

By contrast, the mean of $g(X)$ is given by

$$\begin{aligned} \mathbf{E}[g(X)] &= \frac{1}{2^m - 1} \sum_{l=1}^m \binom{m}{l} \phi \left(-\frac{1}{2} N^{\frac{1}{2}} 2\epsilon l^{\frac{1}{2}} \right) \\ &= \frac{1}{2^m - 1} \sum_{l=1}^m \binom{m}{l} \phi \left(- (N\epsilon^2)^{\frac{1}{2}} l^{\frac{1}{2}} \right). \end{aligned}$$

We now consider the values of these two expressions for a particular example. We suppose that there are $m = 8$ linear approximations, so the capacity $\bar{c}^2 = 8\epsilon^2$. We further suppose that we have $N = 2\epsilon^{-2}$ plaintext-ciphertext pairs, so $N\epsilon^2 = 2$. In this case we have

$$g(\mathbf{E}[X]) = 0.0023, \text{ whereas } \mathbf{E}[g(X)] = 0.0074.$$

For this example, we have $g(\mathbf{E}[X]) > 3\mathbf{E}[g(X)]$, so illustrating Jensen's inequality. However, despite Jensen's inequality, it is essentially asserted by the "proof" of *Corollary 1* that $g(\mathbf{E}[X]) = \mathbf{E}[g(X)]$. (We note that the function $f(x) = g(-x)$ is erroneously used in this "proof" instead of $g(x)$.) This example shows that this assertion, which is the basis of the "proof" of *Corollary 1*, is simply wrong. As with the discussion by [1] of probabilities for dependent data masks [5], the given "proof" of *Corollary 1* by [1] is not correct.

We now calculate the various values given for the gain in this situation, so

$$\gamma = H_8(g(\mathbf{E}[X])) = 5.75 \text{ and } \hat{\gamma} = H_8(\mathbf{E}[g(X)]) = 6.88.$$

Furthermore, a direct calculation gives $\tilde{\gamma} = 6.87$, so $\hat{\gamma}$ is obviously a very good approximation of $\tilde{\gamma}$. In this situation, *Corollary 1* overestimates the gain as given by *Theorem 1* by over one bit in six.

6 Conclusions

We have shown that the value for the gain given by *Corollary 1* of [1] is not reliable, and is in general a large overestimate of the value of the gain given

by *Theorem 1*. Furthermore, the “proof” given of *Corollary 1* simply ignores Jensen’s inequality, a fundamental result in probability and theoretical statistics. Any result based on this value for the gain given by *Corollary 1*, such as the theoretical data requirements for such a linear cryptanalysis, is therefore highly questionable.

References

1. A. Biryukov, C. De Cannière, and M. Quiquater. On Multiple Approximations. In M. Franklin, editor, *Advances in Cryptology – CRYPTO 04*, volume 3152 of *LNCS*, pages 1–22. Springer–Verlag, 2004.
2. B.S. Kaliski and M.J.B. Robshaw. Linear Cryptanalysis Using Multiple Approximations. In Y. Desmedt, editor, *Advances in Cryptology – CRYPTO 94*, volume 839 of *LNCS*, pages 26–39. Springer–Verlag, 1994.
3. B.S. Kaliski and M.J.B. Robshaw. Linear Cryptanalysis Using Multiple Approximations and FEAL. In B. Preneel, editor, *Fast Software Encryption 1994*, volume 1008 of *LNCS*, pages 249–264. Springer–Verlag, 1995.
4. M. Matsui. Linear Cryptanalysis for the DES Cipher. In T. Helleseeth, editor, *Advances in Cryptology – EUROCRYPT 1993*, volume 765 of *LNCS*, pages 386–397. Springer–Verlag, 1993.
5. S. Murphy. The Independence of Linear Approximations in Symmetric Cryptology. *IEEE Transactions on Information Theory*, 52:5510–5518, 2006.
6. S.D. Silvey. *Statistical Inference*. Chapman and Hall, 1975.

A Value for the Gain given by *Theorem 1*

The value γ given for the gain given by *Theorem 1* of [1] is

$$\gamma = -\log_2 \left[2 \frac{1}{|\mathcal{Z}|} \sum_{\mathbf{z} \in \mathcal{Z}^*} \phi \left(-\frac{1}{2} N^{\frac{1}{2}} |\mathbf{c}_{\mathbf{z}} - \mathbf{c}_{\mathbf{z}^*}| \right) + \frac{1}{|\mathcal{Z}|} \right].$$

Thus we have

$$\gamma = -\log_2 \left[2 (1 - 2^{-m}) \frac{1}{2^m - 1} \sum_{\mathbf{z} \neq 0} g (|\mathbf{c}_{\mathbf{z}} - \mathbf{c}_0|^2) + 2^{-m} \right].$$

However, the mean value of $g(X)$ is given by

$$\mathbf{E}[g(X)] = \frac{1}{2^m - 1} \sum_{\mathbf{z} \neq 0} g (|\mathbf{c}_{\mathbf{z}} - \mathbf{c}_0|^2),$$

so we have shown that the value γ given for the gain by *Theorem 1* is given by

$$\gamma = -\log_2 [2 (1 - 2^{-m}) \mathbf{E}[g(X)] + 2^{-m}] = H_m(\mathbf{E}[g(X)]).$$

B Value for the Gain given by *Corollary 1*

The value $\tilde{\gamma}$ given for the gain by *Corollary 1* of [1] is

$$\begin{aligned}\tilde{\gamma} &= -\log_2 \left[2^{\frac{|\mathcal{Z}|-1}{|\mathcal{Z}|}} \phi \left(- \left(\frac{1}{2} N \bar{c}^2 \right)^{\frac{1}{2}} \right) + \frac{1}{|\mathcal{Z}|} \right] \\ &= H_m \left(\phi \left(- \frac{1}{2} (2N\bar{c}^2)^{\frac{1}{2}} \right) \right) = H_m (g(2\bar{c}^2)).\end{aligned}$$

We express this quantity in terms of the mean of X , which is given by

$$\mathbf{E}[X] = \frac{1}{|\mathcal{Z}^*|} \sum_{\mathbf{z} \neq \mathbf{z}^*} |\mathbf{c}_{\mathbf{z}} - \mathbf{c}_{\mathbf{z}^*}|^2 = (2^m - 1)^{-1} \sum_{\mathbf{z} \neq \mathbf{0}} |\mathbf{c}_{\mathbf{z}} - \mathbf{c}_0|^2.$$

However, $|\mathbf{c}_{\mathbf{z}} - \mathbf{c}_0|^2 = 4 \sum_{i=1}^m z_i^2 \epsilon_i^2$ (Section 1), so we have

$$\mathbf{E}[X] = (2^m - 1)^{-1} \sum_{\mathbf{z} \neq \mathbf{0}} \sum_{i=1}^m 4z_i^2 \epsilon_i^2 = \frac{4}{2^m - 1} \sum_{i=1}^m \epsilon_i^2 \sum_{\mathbf{z} \neq \mathbf{0}} z_i^2 = \frac{4 \cdot 2^{m-1}}{2^m - 1} \bar{c}^2,$$

as the two summations in the above expression can be evaluated to give

$$\sum_{\mathbf{z} \neq \mathbf{0}} z_i^2 = \sum_{\mathbf{z} \neq \mathbf{0}} z_i = 2^{m-1} \text{ and } \sum_{i=1}^m \epsilon_i^2 = \bar{c}^2.$$

Thus we can give the capacity in terms of the mean of X as

$$\bar{c}^2 = \frac{1}{2} (1 - 2^{-m}) \mathbf{E}[X],$$

so we can obtain

$$\phi \left(- \left(\frac{1}{2} N \bar{c}^2 \right)^{\frac{1}{2}} \right) = \phi \left(- \frac{1}{2} N^{\frac{1}{2}} \left((1 - 2^{-m}) \mathbf{E}[X] \right)^{\frac{1}{2}} \right) = g \left((1 - 2^{-m}) \mathbf{E}[X] \right).$$

This means we can express the value $\tilde{\gamma}$ given for the gain by *Corollary 1* as

$$\tilde{\gamma} = H_m (g((1 - 2^{-m}) \mathbf{E}[X])).$$

If we now define the value

$$\hat{\gamma} = H_m (g(\mathbf{E}[X])),$$

then clearly $\hat{\gamma}$ is a very good approximation of $\tilde{\gamma}$ when m is moderately large.