Properties of Linear Approximation Tables

Luke O'Connor *,1,2

[1] Distributed Systems Technology Centre (DSTC), Brisbane, Australia
[2] Information Security Research Centre, Queensland University of Technology
GPO Box 2434, Brisbane Q 4001, Australia

Email: oconnor@dstc.edu.au

Abstract. Linear cryptanalysis is an attack that derives a linear approximation between bits of the plaintext, ciphertext and key. This global approximation is constructed from the linear approximation tables of the nonlinear mappings used by the cipher, usually the S-boxes, as in the case of DES. In this paper we will describe the distribution of these tables for bijective mappings (permutations), concentrating on the expected value of the largest entry, and use our results to construct Feistel ciphers provably resistant to linear cryptanalysis.

1 Introduction

Linear cryptanalysis [11, 10] is a recently proposed attack due to M. Matsui. When successful, the attack recovers information about the secret key K used by approximating several nonlinear components of the cipher. For DES, the S-boxes can be approximated by deriving linear relations between the inputs and outputs to each S-box, where each relation is true with some probability p_i. We will call each such approximation a *linearization*. Matsui has shown that it is possible to derive linearizations to the S-boxes of DES at various rounds such that when these linearizations are added modulo 2, the remaining linearization involves bits of the key, plaintext and ciphertext only. The probability p of this 'global' linearization being correct is determined directly from the probabilities p_i of the 'local' linearizations when it is assumed that the subkeys are independent. One bit of information concerning the key can be recovered using maximum likelihood estimation when approximately $|p - \frac{1}{2}|^{-2}$ plaintext-ciphertext pairs are known. The quantity $|p - \frac{1}{2}|^{-2}$ is referred to as the (data) complexity of the attack. The attack can be modified to obtain more information about the key.

There have been several responses to the introduction of linear cryptanalysis. Kim *et al.* [8] have given a list of conditions for DES-like S-boxes to satisfy where

* The work reported in this paper has been funded in part by the Cooperative Research Centres program through the Department of the Prime Minister and Cabinet of Australia.

the probability p of the best global approximation based on these S-boxes will have the property that $|p - \frac{1}{2}| > 2^{28}$, implying that the complexity of the attack exceeds the cost of exhaustive key search. More general ciphers resistant to linear cryptanalysis have been proposed by Knudsen [9] and Heys and Tavares [7]. Also, several similarities between linear cryptanalysis and the more familiar differential cryptanalysis [2] have been noted [1, 3, 9].

The main result of this paper is to derive an upper bound on the largest entry in the linear approximation table of a bijective mapping, which when combined with a bound on the number of rounds that must be linearized, yields a lower bound on the complexity of linear cryptanalysis.

2 Linear approximation tables

Let $\pi : Z_2^n \rightarrow Z_2^n$ be a bijective n-bit mapping, and let S_{2^n} be the set of all such mappings, known as the symmetric group. For an n-bit vector $X \in Z_2^n$ let $X[i]$ denote the ith bit of X. The linear approximation table for π, denoted LAT_π, is a $2^n \times 2^n$ table such that

$$LAT_\pi(\alpha, \beta) \stackrel{\text{def}}{=} \#\left\{ X \;\middle|\; X \in Z_2^n,\ \bigoplus_{i=1}^{n} X[i] \cdot \alpha[i] = \bigoplus_{i=1}^{n} \pi(X)[i] \cdot \beta[i] \right\} \quad (1)$$

where $\alpha, \beta \in Z_2^n$ and '$\cdot$' denotes bitwise logical AND. Thus $LAT_\pi(\alpha, \beta)$ gives the number of equal parity checks between a linear combination of the input bits (specified by α) and a linear combination of the output bits (specified by β).

Theorem 1. Let $\lambda(\alpha, \beta)$ be a random variable describing $LAT_\pi(\alpha, \beta)$ when π is selected uniformly from S_{2^n}, and α, β are nonzero. Then $\lambda(\alpha, \beta)$ only assumes even values and

$$\Pr(\lambda(\alpha, \beta) = 2k) = \frac{(2^{n-1}!)^2}{2^n!} \cdot \binom{2^{n-1}}{k}^2. \quad (2)$$

for $0 \leq k \leq 2^{n-1}$. □

In linear cryptanalysis we are interested in those entries in the linear approximation table that differ from 2^{n-1} as this represents the correlation between linear combinations of the inputs and outputs. Matsui [10] calls this the *effectiveness* of the linearization. For this reason we define the 'normalized' linear approximation table, denoted by LAT_π^*, as

$$LAT_\pi^*(\alpha, \beta) = |LAT_\pi(\alpha, \beta) - 2^{n-1}|. \quad (3)$$

The distribution of $LAT_\pi^*(\alpha, \beta)$ follows directly from Theorem 1.

Corollary 2.1 Let $\lambda^*(\alpha, \beta)$ be a random variable describing $LAT_\pi^*(\alpha, \beta)$ when π is selected uniformly from S_{2^n}, and α, β are nonzero. Then $\lambda^*(\alpha, \beta)$ only assumes even values and

$$\Pr(\lambda^*(\alpha, \beta) = 2k) = \frac{(2 - [k = 0]) \cdot (2^{n-1}!)^2}{2^n!} \cdot \binom{2^{n-1}}{2^{n-2} + k}^2. \quad (4)$$

for $0 \leq k \leq 2^{n-2}$, where $[k = 0]$ evaluates to zero or one. □

Using Stirling's approximation, it can be shown that the expected number of zero entries in the table is $4/\sqrt{2\pi \cdot 2^n}$, which is tending to zero with n. However, our main goal is to determine a bound on the largest entry in LAT_π^* which is useful in the design of ciphers that are to be resistant to linear cryptanalysis. To this end let $\lambda(\pi)$ be the largest entry in LAT_π^* for the mapping π taken over all nontrivial α, β,

$$\lambda(\pi) \stackrel{\text{def}}{=} \max_{\alpha,\beta \neq 0} LAT_\pi^*(\alpha, \beta). \tag{5}$$

In the next section we derive an upper bound on $\lambda(\pi)$.

3 An upper bound

Let $\mathbf{E}[\lambda(\pi, 2k)]$ denote the expected number of entries in LAT_π^* of size $2k$. Consider the following bound on $\lambda(\pi)$,

$$\Pr(\lambda(\pi) = 2k) \leq \Pr(LAT_\pi^* \text{ has at least one entry of size } 2k) < \mathbf{E}[\lambda(\pi, 2k)] \tag{6}$$

which is valid since

$$\mathbf{E}[\lambda(\pi, 2k)] = \sum_{t \geq 0} t \cdot \Pr(LAT_\pi^* \text{ has } t \text{ entries of size } 2k). \tag{7}$$

If $\mathbf{E}[\lambda(\pi, 2k)]$ is tending rapidly to zero as a function of k, then we are likely to obtain a useful bound on $\pi(\lambda)$. From Corollary 2.1, we can derive $\mathbf{E}[\lambda(\pi, 2k)]$, $k > 0$ as

$$\mathbf{E}[\lambda(\pi, 2k)] = \frac{2 \cdot (2^n - 1)^2 \cdot (2^{n-1}!)^2}{2^n!} \cdot \binom{2^{n-1}}{2^{n-2} + k}^2. \tag{8}$$

In the next theorem we derive an approximation for the tail of the $\mathbf{E}[\lambda(\pi, 2k)]$ distribution using a well-known bound on the sum of consecutive binomial coefficients.

Theorem 2. For $t > 0$,

$$\sum_{k=2^{n-2}+t}^{2^{n-1}} \mathbf{E}[\lambda(\pi, 2k)] < \sqrt{\frac{\pi}{2}} \cdot 2^{2^n H(m) - 2^n + 5n/2} \tag{9}$$

where $m = (2^{n-2} - t)/2^{n-1}$ and $H(x) = -x \log x - (1 - x) \log(1 - x)$. □

From Theorem 2 we are now able to derive a bound on the probability that $\lambda(\pi)$ is at most $2t$ since

$$\Pr(\lambda(\pi) \leq 2t) > 1 - \sum_{k=2^{n-2}+t}^{2^{n-1}} \mathbf{E}[\lambda(\pi, 2k)]. \tag{10}$$

In particular, we will determine the smallest value of t (according to our bounds) for which $\Pr(\lambda(\pi) \leq 2t) > \frac{1}{2}$, which we will denote as t_n. Here t_n is referred to as the median of the distribution [6]. The practical implication of t_n is that when a bijective n-bit mapping π is selected uniformly, with odds better than 50/50 the mappings will have $\lambda(\pi) \leq 2t_n$.

Our results are shown in Table 1. The second column shows the actual tail computation for $\mathbf{E}[\lambda(\pi, 2k)]$ using (8), while the third column shows the tail computation using the approximation of (9). The last column shows the results of generating random bijective mappings and determining t_n from this sample. That is, if there were k mappings generated, t_n is determined to be the minimum value for which at least half the mappings had the largest entry bounded by t_n. The sample k actually used was relatively small (several hundred) because of the time required to determine t_n. For example, it requires half an hour of clock time to determine t_{10} for one 10-bit mapping.

n	t_n via (8)	t_n via (9)	t_n via experiment
6	9	10	8
7	13	15	12
8	19	22	18
9	28	33	26
10	41	49	40
11	61	72	59
12	90	106	88

Table 1. Upper bounds on t_n where $\Pr(\lambda(\pi) \leq 2t_n) > \frac{1}{2}$.

Note that the bound on t_n derived from (9) is consistently higher than the bound derived from (8). Of course this is expected as various approximations are used to derive the bound in (9). The advantage however is that the expression in (9) is easy to evaluate and can give useful information about the best approximation for large mappings. On the other hand, the expression in (8) can only be evaluated for relatively small n. For example, using (9), we have determined that the largest entry in the linear approximation table of a 64-bit mapping is at most 16057555882 for more than half the possible such mappings, giving a best possible linear approximation of

$$|p - \frac{1}{2}| \leq \frac{16057555882}{2^{64}} = 0.87048 \times 10^{-9} \approx 2^{-30}. \tag{11}$$

We may then assume that the probability of the best (global) linear approximations for 64-bit mappings such as DES and FEAL would be close to 2^{-30} *if the mappings were random and not derived from an iterative process*. However, using the fact that the nonlinear components of the round function are fixed, Matsui has derived an approximation for 16-round DES that has $|p - \frac{1}{2}| = 1.19 \times 2^{-21}$.

4 Conclusion

The success of linear cryptanalysis depends on the nonlinearity of the round function **F**, which for ciphers such as DES reduces to the nonlinearity of the S-boxes. It is clear that if this nonlinearity can be made sufficiently large then the complexity of the attack will exceed the cost of exhaustive key search. Our approach has been to bound the largest value in the linear approximation table of a bijective mapping, which can now be used to construct ciphers resistant to linear cryptanalysis using similar methods to those employed by Knudsen [9] and Heys and Tavares [7]. Similar observations have been used by O'Connor [13] to show that sufficiently large S-boxes will also defeat differential cryptanalysis, as the largest value in the XOR table can be bounded asymptotically.

Acknowledgements

I would like to thank Kaisa Nyberg for for her comments on the original version of this manuscript.

References

1. E. Biham. On Matsui's Linear Cryptanalysis. *to appear, proceedings of EUROCRYPT 94, Perugia, Italy*, 1994.
2. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.
3. F. Chabaud and S. Vandenay. Links between differential and linear cryptanalysis. *to appear, proceedings of EUROCRYPT 94, Perugia, Italy*, 1994.
4. H. Feistel. Cryptography and computer privacy. *Scientific American*, 228(5):15–23, 1973.
5. H. Feistel, W. A. Notz, and J. Lynn Smith. Some cryptographic techniques for machine-to-machine data communications. *proceedings of the IEEE*, 63(11):1545–1554, 1975.
6. W. Feller. *An Introduction to Probability Theory and its Applications.* New York: Wiley, 3rd edition, Volume 1, 1968.
7. H. M. Heys and S. E. Tavares. Substitution-permutation networks resistent to differential and linear cryptanalysis. submitted to the Journal of Crytology.
8. K. Kim, S. Lee, S. Park, and D. Lee. DES can be immune to linear cryptanalysis. *proceedings of the Workshop on Selected Areas in Cryptography, Kingston, Canada, May 1994*, pages 70–81, 1994.
9. L. R. Knudsen. Practically secure Feistel ciphers. *proceedings of Fast Software Encryption, Cambridge Security Workshop, Lecture Notes in Conputer Science, vol. 809, 1994*, pages 211–221, 1994.
10. M. Matsui. Linear cryptanalysis of DES cipher (I). (version 1.03) private communication.
11. M. Matsui. Linear cryptanalysis method for DES cipher. *Advances in Cryptology, EUROCRYPT 93, Lecture Notes in Computer Science, vol. 65, T. Helleseth ed., Springer-Verlag*, pages 386–397, 1994.

12. W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. *Advances in Cryptology, EUROCRYPT 89, Lecture Notes in. Computer Science, vol. 434, J.-J. Quisquater, J. Vandewalle eds., Springer-Verlag*, pages 549–562, 1990.
13. L. J. O'Connor. On the distribution of characteristics in bijective mappings. *Advances in Cryptology, EUROCRYPT 93, Lecture Notes in Computer Science, vol. 765, T. Helleseth ed., Springer-Verlag*, pages 360–370, 1994.
14. J. Pieprzyk, C. Charnes, and Seberry J. Linear approximation versus nonlinearity. *proceedings of the Workshop on Selected Areas in Cryptography, Kingston, Canada, May 1994*, pages 82–89, 1994.