SBoxLin.tex, 11.02.03

Generalised S-Box Nonlinearity

NES/DOC/UIB/WP5/020/A

Matthew G. Parker [1] [2]
Institute for Informatics,
University of Bergen, Norway

Abstract

In this paper the (effective) bias of certain generalised linear approximations to the S-box are considered. Whereas, in the literature, the cryptanalyst typically restricts this search to linear approximations over Z_2, we here consider linear approximations over Z_4 and, more generally still, consider approximations which are linear in the sense that they can be completely factorised into the tensor product of length-two vectors. Consequently, significantly higher biases can be found in comparison to Z_2-linear approximations.

I. Introduction and Basic Theory

Linear cryptanalysis of a binary block cipher, as conceived by Matsui [3], attempts to recover key bits by approximating core rounds of the block cipher by a series of Z_2-linear expressions, and then concatenating these linear approximations. The block cipher to be approximated is parameterised by a secret key which is typically (although not always) added into the cipher by means of XOR. If this is the case then the binary linear approximation to the block cipher core is key-invariant to within a global constant. If some subset of the key bits are not added using XOR, then this subset is often taken to be fixed, so that a complete linear approximation can be established that holds for a subset of all possible key configurations. The linear approximation can be used to relate certain input bits to certain output bits of the core rounds of a block cipher, where the approximation holds with probability $\frac{1}{2} \pm b$, where b is called the bias. The cryptanalyst can then use a set of known plaintext/ciphertext pairs to ascertain key bits of the first and/or last rounds of the cipher (which are outside the core), by guessing the key bits, and then checking whether the relevant output and/or input bits of the core rounds are correct with probability $\frac{1}{2}$ or with probability $\frac{1}{2} \pm b$. If it is the latter then, given a large enough set of plaintext/ciphertext pairs, the key guess is probably correct.

To counteract this form of attack many modern block ciphers, including the AES (Rijndael), make it as difficult as possible to approximate the constituent functions of the Substitution-Boxes (S-boxes) within the cipher by Z_2-linear functions. Let $\mathbf{x}$ and $\mathbf{y}$ be short for the $2n$ input and output boolean variables, $(x_0, x_1, \ldots, x_{n-1})$ and $(y_0, y_1, \ldots, y_{n-1})$, that define the n input and n output bits of an $n \times n$ S-box, respectively. Also let f be any Z_2-linear combination of the variables of $\mathbf{x}$, g be any Z_2-linear combination of the variables of $\mathbf{y}$, and let $\mathbf{A}$ be a subset of the size 2^n set $\mathbf{S}$ where $\mathbf{S}$ contains all 2^n $(\mathbf{x}, \mathbf{y})$ pairs that define the given S-box. Then, to be precise, a $n \times n$ S-box is considered to be

[1] The work described in this paper has been supported by the Commission of the European Communities through the IST program under contract IST-1999-12324

optimally resistant to Linear Cryptanalysis if the following conditions hold,

$$\begin{array}{ll} f(\mathbf{x}) = g(\mathbf{y}) & \forall(\mathbf{x},\mathbf{y}) \in \mathbf{A} \\ ||\mathbf{A}| - 2^{n-1}| \leq 2^{\frac{n}{2}} & n \text{ even} \\ ||\mathbf{A}| - 2^{n-1}| \leq 2^{\frac{n-1}{2}} & n \text{ odd} \end{array} \tag{1}$$

although, for the n odd case, it is an open problem to prove that the right-hand side is really the lowest possible value. The S-box, an n-bit to n-bit invertible mapping, is completely described by the set of $(\mathbf{x}, \mathbf{y})$ pairs, $\mathbf{S}$, and $\mathbf{A}$ contains the complete subset of the elements of $\mathbf{S}$ for which $f(\mathbf{x}) = g(\mathbf{y})$ holds.

Typically approximation is done, for binary ciphers, by finding the nearest Z_2-linear approximation to selected nonlinear segments of the cipher. These approximations are then pieced together and the biases of each segment simply multiplied together under the important assumption of virtual pairwise independence of any two segments whose outputs are combined. The largest Z_2-linear approximations can be found via spectral analysis with respect to (wrt) the Walsh-Hadamard Transform (WHT). This paper shows how one can, more generally, choose to approximate the constituent functions of a binary S-box by **any** linear function over **any** weighted alphabet. However this does **not** imply that it is straightforward to piece these generalised linear approximations together in exactly the same way as for standard linear cryptanalysis. In other words, for a block cipher which adds in the key using XOR, the way in which the generalised linear approximations are concatenated is, in general, **key-dependent**. Techniques to piece together these generalised linear approximations are left to future research. In this paper we simply investigate the biases of certain generalised linear approximations to certain S-boxes. Crucial to our approach is the use of the tensor product to define a very general form of linearity. Specifically we state that *a tensor-linear sequence is any normalised length N sequence (with truth-table outputs ordered lexicographically) which can be fully tensor-decomposed according to the factors of N.* For example, when $N = 2^n$, then a tensor-linear length-N sequence can be written in the form,

$$(a_0, b_0) \otimes (a_1, b_1) \otimes \ldots \otimes (a_{n-1}, b_{n-1})$$

We observe that the rows of the $N \times N$ Walsh-Hadamard Transform (WHT) encompass all tensor-linear sequences (to within a global multiplicative offset) which can be written in the form,

$$(\pm 1, \pm 1) \otimes (\pm 1, \pm 1) \otimes \ldots \otimes (\pm 1, \pm 1)$$

To be explicit here, as an example, is the 4×4 WHT matrix:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} (1,1) & \otimes & (1,1) \\ (1,-1) & \otimes & (1,1) \\ (1,1) & \otimes & (1,-1) \\ (1,-1) & \otimes & (1,-1) \end{pmatrix}$$

Thus when we are correlating our n-variable binary S-Box function, $f(\mathbf{x})$, with rows of the WHT, we are in fact correlating with bipolar tensor-linear sequences. Note that, in the cryptographic literature, the more conventional way to write the action of the WHT on $f(\mathbf{x})$ is as follows:

$$F_{\mathbf{k}} = 2^{-n} \sum_{\mathbf{x} \in Z_2^n} (-1)^{f(\mathbf{x}) + \mathbf{x} \cdot \mathbf{k}} \tag{2}$$

where $\mathbf{k} \in Z_2^n$, $f : Z_2^n \rightarrow Z_2$, and $\mathbf{x} \cdot \mathbf{k}$ is the inner product of $\mathbf{x}$ and $\mathbf{k}$. We define the Peak-to-Average Power Ratio (PAR) with respect to (wrt) the WHT as,

$$\mathrm{PAR}(f) = 2^n \max_{\forall \mathbf{k}}(|F_{\mathbf{k}}|)^2 \tag{3}$$

In terms of the more familiar cryptographic measure of bias,

$$\mathrm{PAR}(f) = \mathrm{bias}^2 \times 2^{n+2} \tag{4}$$

The value of PAR ranges from 1.0 for a completely flat spectrum, to 2^n for a linear function. PAR with respect to the WHT measures the goodness of the highest possible Z_2-linear approximation to f. The higher PAR is, the better the approximation.

More generally, we can define the PAR of f with respect to any normalised set of transforms, $\mathbf{T}$, which produces sets of spectra, $\{F_{\mathbf{k}}\}$, using (3), so we can, instead, maximise PAR wrt the set of transforms, $\mathbf{T}$. We now have,

$$\mathrm{PAR}(f) = 2^n \max_{\forall \mathbf{k}, \forall \mathbf{U} \in \mathbf{T}}(|F_{\mathbf{k}}|)^2 \tag{5}$$

Once again, the higher PAR is, the better the approximation. In this paper we examine aspects of the nonlinearity of certain S-boxes that have been proposed in the cryptographic literature, and this nonlinearity is quantified in terms of PAR of the spectra of the constituent boolean functions with respect to selected sets, $\mathbf{T}$, of complex unitary transforms with linear rows [3]. From the evaluations of this paper one can conclude that, although (by design) it is hard to find good Z_2-linear approximations to the S-boxes of modern block ciphers, much better generalised linear approximations exist, and these improved approximations may perhaps be useful in the context of novel cryptanalysis of the block cipher in which the S-Box is used.

II. S-Box Nonlinearity

This section details the best (highest) and worst (lowest) Peak-to-Average power ratio (PAR) of various S-boxes in the literature with respect to various sets of Linear Unitary Transforms (LUTs). Let our S-Box have n binary inputs, $\mathbf{x} = (x_0, x_1, \ldots, x_{n-1})$, and m binary outputs, $\mathbf{y} = (y_0, y_1, \ldots, y_{m-1})$. Then we can describe our S-Box by the boolean functions,

$$y_i = f_i(\mathbf{x}), \qquad 0 \leq i < m, \qquad f_i : Z_2^n \rightarrow Z_2$$

[3]Linearity here refers to 'tensor-linearity', where each length 2^n row of the 'Linear Unitary Transform' (LUT) matrix can be written as a tensor-product of length 2 complex vectors [7].

Then we wish to find the largest (and smallest) generalised linear approximations, taken over all functions, $f : Z_2^n \rightarrow Z_2$, of the form,

$$f(\mathbf{x}) = \sum_{i=0}^{m-1} c_i f_i(\mathbf{x}), \qquad c_i \in Z_2 \tag{6}$$

and we will use PAR to quantify the goodness of the linear approximation - a low PAR means a bad approximation, and a high PAR means a good approximation.

Note that, although the metric of nonlinearity, γ, is strictly only defined wrt the Walsh-Hadamard Transform (WHT), we can relate a generalised form of γ to the PAR of a given function, $f : Z_2^n \rightarrow Z_2$, by,

$$\gamma(f) = 2^{\frac{n}{2}-1}(2^{\frac{n}{2}} - \sqrt{\text{PAR}(f)}) \tag{7}$$

However, for brevity, the results of this paper will only be quoted in terms of PAR.

We will compute largest and smallest PARs of f wrt the WHT, the **HN** transform set, the **HI** transform set, and the **HIN** transform set, where the WHT transform matrix comprises tensor products of $\mathbf{H} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, the **HN** transform matrix comprises all 2^n combinations of tensor products of **H** and $\mathbf{N} = \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}$, the **HI** transform matrix comprises all 2^n combinations of tensor products of **H** and $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and the **HIN** transform matrix comprises all 3^n combinations of tensor products of **H**, **I**, and **N**. To be precise,the largest PAR of f wrt the transform set, $\mathbf{T}$, is computed by correlating $(-1)^f$ with all rows of all of the transforms in $\mathbf{T}$, and finding the highest magnitude correlation. The largest PAR is then obtained by squaring this highest magnitude and normalising appropriately. In contrast, the smallest PAR of f wrt the transform set, $\mathbf{T}$, is computed by first finding the maximum correlation magnitudes taken over all rows of $\mathbf{U}$, where $\mathbf{U} \in \mathbf{T}$, one maximum per matrix, $\mathbf{U}$. Then the smallest PAR is obtained by selecting the smallest of these maximum magnitudes, then squaring and normalising this value. Unless otherwise indicated, throughout this paper the 'largest PAR' of f wrt the transform set, $\mathbf{T}$, will be simply referred to as the 'PAR' of f.

There are an infinite number of Linear Unitary Transforms (LUTs) so the above tensor combinations of **H**, **I**, and **N** acting on f only give a lower bound on the actual PAR of f with respect to all LUTs, but we will show that even this small subset of transforms can identify higher biases than are achievable with respect to the WHT. For small S-Boxes we will also compute the **HNF** transform of f, where **HNF** comprises all 4^n combinations of tensor products of **H**, **N**, $\mathbf{F_0}$, and $\mathbf{F_1}$, where, $\mathbf{F_0} = \begin{pmatrix} 1 & \omega \\ 1 & -\omega \end{pmatrix}$ and $\mathbf{F_1} = \begin{pmatrix} 1 & \omega^3 \\ 1 & -\omega^3 \end{pmatrix}$, where ω is the eighth complex root of 1, such that $\omega^2 = i$. The **HNF** finds Z_8-linear approximations to the S-box functions.

In the following we use the above transform matrices to examine every possible generalised linear relationship between a subset of inputs and outputs of the specific S-box. This is accomplished by applying the transform matrices to all possible f, where f is given by (6). The largest and smallest biases are obtained from the output spectra from these transforms and recorded in the tables in terms of PAR. One should note that the spectra examined by the **HI** set of transforms includes that covered by the WHT. Similary WHT is contained within **HN**, and **HI** and **HN** are both contained in **HIN**, and both contained in **HNF**. In general, if the set of transforms, **T**, is contained within the set of transforms, **U**, then the PAR of a function with respect to **T** acts as a lower bound on the PAR of the function with respect to **U**.

Finally it is important to stress that we are only, in this paper, looking at the transform spectra of Z_2-linear combinations of the outputs of the constituent S-box functions, as indicated by (6). More generally still we could examine generalised linear approximations of **generalised linear** combinations, f, of the S-box **outputs**. We do not consider this further generalisation in this paper, although preliminary investigations indicate the existence of extremely high biases for every S-box examined.

A. DES - 6 Input, 4 Output

DES is the well-known Data Encryption Standard [4].

	Largest PARs				
	WHT	**HN**	**HI**	**HIN**	**HNF**
S-Box 1	20.25	20.25	24.5	24.5	20.25
S-Box 2	16.0	16.0	18.0	18.0	16.00
S-Box 3	16.0	16.0	16.0	21.125	16.00
S-Box 4	16.0	18.0	18.0	18.0	19.90
S-Box 5	25.0	25.0	25.0	25.0	25.00
S-Box 6	12.25	13.625	16.0	16.0	14.45
S-Box 7	20.25	20.25	21.125	21.125	20.25
S-Box 8	16.0	16.0	21.125	21.125	16.42

	Smallest PARs				
	WHT	**HN**	**HI**	**HIN**	**HNF**
S-Box 1	4.0	6.25	8.0	9.0	7.25
S-Box 2	4.0	6.25	8.0	9.0	6.58
S-Box 3	6.25	6.625	9.0	9.0	7.66
S-Box 4	6.25	8.5	9.0	9.0	9.27
S-Box 5	4.0	5.125	9.0	9.0	5.87
S-Box 6	4.0	6.25	9.0	9.0	7.54
S-Box 7	4.0	5.625	8.0	8.0	7.29
S-Box 8	6.25	6.25	9.0	9.0	7.06

For DES, the PARs with respect to the **HN** transform are not much better then those with respect to the WHT. However, improved correlations are found by considering **HI** and **HIN** transforms. The resistance to linear attacks of the DES S-boxes is already quite weak with respect to WHT, whereas for Rijndael and Serpent, say, the S-boxes have been designed to be 'optimally' nonlinear with respect to WHT. Note that the **HNF** transform improves on the **HN** transform, but usually the **HI** and **HIN** transforms find stronger biases than the **HNF** transform, apart from S-Box 4.

B. Serpent - 4 Input, 4 Output

	Largest PARs				
	WHT	**HN**	**HI**	**HIN**	**HNF**
S-Box 0	4.0	8.0	8.0	8.0	8.0
S-Box 1	4.0	8.0	8.0	8.0	8.0
S-Box 2	4.0	8.0	8.0	8.0	8.0
S-Box 3	4.0	5.0	8.0	8.0	5.83
S-Box 4	4.0	8.0	8.0	8.0	8.0
S-Box 5	4.0	8.0	8.0	8.0	8.0
S-Box 6	4.0	8.0	8.0	8.0	8.0
S-Box 7	4.0	5.0	8.0	8.0	6.83

	Smallest PARs				
	WHT	**HN**	**HI**	**HIN**	**HNF**
S-Box 0	4.0	4.0	4.0	4.5	4.0
S-Box 1	4.0	4.0	4.0	4.5	4.0
S-Box 2	4.0	4.0	4.5	4.5	4.0
S-Box 3	4.0	4.0	4.5	4.5	4.0
S-Box 4	4.0	4.0	4.0	4.5	4.0
S-Box 5	4.0	4.0	4.0	4.5	4.0
S-Box 6	4.0	4.0	4.0	4.5	4.0
S-Box 7	4.0	4.0	4.5	4.5	4.0

Note that all Serpent S-boxes have significantly closer linear approximations with respect to **HI** or **HIN** transforms. A PAR of 4.0 implies a nonlinearity of $\gamma = 4.0$, whereas a PAR of 8.0 implies a nonlinearity of $\gamma = 2.34$.

C. Rijndael - 8 Input, 8 Output

	Largest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	4.0	14.125	18.0	18.0

	Smallest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	4.0	7.031	12.25	12.25

A PAR of 4.0 is equivalent to a nonlinearity of $\gamma = 112$, and a PAR of 18.0 is equivalent to a nonlinearity of $\gamma = 94.06$.

D. Khazad - 8 Input, 8 Output

	Largest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	16.0	16.0	22.78	22.78

	Smallest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	5.06	7.53	11.28	12.16

The biases for Khazad are noticeably stronger than for, say, Rijndael. Khazad is not optimised for the WHT transform, and the **HN** transform does not improve much on the WHT. However, both **HI** and **HIN** transforms do uncover significantly stronger biases than the WHT. There is an interesting trade-off here between the Khazad S-box and the Rijndael S-box. The Khazad S-box is much simpler to implement, as it is built out of a combination of 4×4 miniboxes. But the price paid for this simplicity is a reduced nonlinearity with respect to the WHT. However, the difference in nonlinearity between Khazad and Rijndael is somewhat moderated when one considers more general linear approximations.

E. Whirlpool - 8 Input, 8 Output

	Largest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	12.25	14.28	22.78	22.78

	Smallest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	4.0	7.56	12.25	12.25

The biases for Whirlpool are noticeably stronger than for, say, Rijndael, and similar to Khazad. Whirlpool is not optimised for the WHT transform, and the **HN** transform does not improve much on the WHT. However, both **HI** and **HIN** transforms do uncover significantly stronger biases than the WHT.

F. MISTY1 - 7/9 Input, 7/9 Output

MISTY1 uses two S-boxes, the first has 7 inputs and 7 outputs, and the second has 9 inputs and 9 outputs. Matsui argues that the reason for choosing odd numbers of binary variables as i/o is because one can find functions with stronger nonlinearity (lowest possible PAR = 2.0 wrt WHT) as compared to functions of even numbers of binary variables (lowest possible PAR = 4.0 wrt WHT). It is therefore of interest to see whether this advantage is carried over to more generalised linear approximations. Here are the results for MISTY1:

SBox S7 - 7 binary inputs, 7 binary outputs

	Largest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	2.0	10.0	16.0	16.0

	Smallest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	2.0	5.0	9.0	9.0

SBox S9 - 9 binary inputs, 9 binary outputs

	Largest PARs			
	WHT	**HN**	**HI**	**HIN**
S-Box	2.0	32.0	32.0	?

	Smallest PARs			
Smallest PARs:	WHT	**HN**	**HI**	**HIN**
S-Box	2.0	4.0	8.0	?

The results for MISTY1 show that, although the S-Boxes exhibit high nonlinearity wrt WHT, they have much lower nonlinearity wrt more general linear approximations. In particular, they can be approximated strongly by rows of the **HI** transform. This is perhaps not surprising as S7 and S9 comprise boolean functions of relatively low degree, and the **HI** transform implicitly involves statistical fixing of bits [6]. In terms of bias, S7 has a bias of $\frac{1}{16}$ wrt WHT and $\frac{1}{\sqrt{8}} = 0.354$ wrt the **HI** transform. Similarly, S9 has a bias of $\frac{1}{32}$ wrt WHT and $\frac{1}{8}$ wrt the **HI** transform. From the (rather little) experience the author has obtained by finding the largest generalised linear approximations to the constituent boolean functions of a 'well-chosen' S-box, it appears that, for a boolean function of n variables, the largest approximation is associated with a maximum PAR $\simeq 2^{\lceil \frac{n}{2} \rceil}$. This, paradoxically, suggests that PAR wrt generalised linear approximation is minimised by choosing n even. But, for the case of MISTY1, n is 7 or 9, and the largest approximations are associated with PARs of $2^{\lceil \frac{7}{2} \rceil} = 16.0$ and $2^{\lceil \frac{9}{2} \rceil} = 32.0$. The results suggest that the argument for higher nonlinearity wrt WHT with n odd is reversed when one considers nonlinearity wrt more general linear approximations. In other words, for an n input, n output S-box, if it is true that the highest possible PAR wrt general linear approximation is $\simeq 2^{\lceil \frac{n}{2} \rceil}$, then n should be chosen to be even. In any case, the largest PARs of 16.0 and 32.0 for S7 and S9 above show a significant increase over the WHT PARs of 2.0, and this improvement is more significant than that achieved for S-boxes with even numbers of input/output variables.

III. Comments

It is clear from the results of this paper that higher biases are possible across state-of-the-art S-boxes by using generalised linear approximations than by using standard Z_2-linear approximations. This raises the following questions:

What is the lowest possible largest PAR of a function, f, with respect to all possible Linear Unitary Transforms, where f is constructed as in (6) from any possible binary S-box with a fixed number of input bits and output bits?

How would one design such an S-box so that any f constructed from the S-box using (6) achieves a minimum largest PAR wrt all Linear Unitary Transforms?

The approximation techniques considered in this paper, and in [8],[6] and [7] may, perhaps, be useful in certain types of cryptanalytic attack. As is known, the output spectra covered by the WHT transform identify tensor-linear approximations which are not key-dependent for block ciphers that add in the round-key using XOR. However, for any other transform, the approximations are key-dependent. But they could still, perhaps, be useful in the context of attacks which make use of approximate Multivariate Generalised Linear system solvers.

In [7] we formalise many of the ideas that have been discussed in this paper, and in [8] and [6]. These tensor-linear approximations essentially utilise in a cryptographic context the Quantum Entanglement metric previously proposed in [5].

References

[1] R.Anderson,E.Biham,L.Knudsen, "Serpent: a Proposal for the Advanced Encryption Standard", *NIST's AES homepage, http://www.nist.gov/aes/*

[2] J.Daemen,V.Rijmen, "The Block Cipher Rijndael," *NIST's AES homepage, http://www.nist.gov/aes/*

[3] M.Matsui, "Linear Cryptanalysis Method for DES Cipher," *Advances in Cryptology - EUROCRYPT '93*, LNCS 765, pp. 386-397, 1994

[4] National Bureau of Standards, "Data Encryption Standards", *FIPS Publication 46*, U.S. Dept. of Commerce, 1977

[5] Parker, M.G., Rijmen, V.: The Quantum Entanglement of Binary and Bipolar Sequences. Short version in **Sequences and Their Applications**, Discrete Mathematics and Theoretical Computer Science Series, Springer, 2001 Long version at *http://xxx.soton.ac.uk/ps/quant-ph/0107106* or *http://www.ii.uib.no/~matthew/* Jun (2001)

[6] M.G.Parker, "Z_2 and Z_4-Linear Subspace Approximations", *NESSIE, NES/DOC/UIB/WP5/019/A*, June, 2002

[7] M.G.Parker, "Complete Linear Cryptanalysis", *NESSIE, NES/DOC/UIB/WP5/021/A*, August, 2002

[8] H.Raddum, M.G.Parker, "Z_4-Linear Cryptanalysis", *NESSIE, NES/DOC/UIB/WP5/018/1*, June, 2002