Structural Attack (and Repair) of Diffused-Input-Blocked-Output White-Box Cryptography

Claude Carlet e,d
Sylvain Guilley a,b,c
Sihem Mesnager e,b

a SECURE-IC THE SECURITY SCIENCE COMPANY
b TELECOM Paris, IP PARIS
c ENS ÉCOLE NORMALE SUPÉRIEURE
d UNIVERSITY OF BERGEN
e UNIVERSITÉ PARIS8 VINCENNES-SAINT-DENIS

Presentor:
Sylvain Guilley, General Manager & CTO, *Secure-IC*

September 2022

IACR Trans. Cryptogr. Hardw. Embed. Syst.2021(4): 57-87 (2021)

State-of-the-art in White-Box Cryptography

White-Box Cryptography (WBC)

- WBC:
  - A hardened version of $m \rightarrow c$=WBC(m)=AES($k^\star$, m), where the secret key $k^\star$ is concealed within the function WBC, which acts as a *public key*
- AES in WBC:
  - Client, can encrypt using c=AES($k^\star$, m)
  - Server, knows the *secret key* $k^\star$, hence can decrypt and ciphertext c
- Use-case:
  - Host Card Emulation (HCE)
  - Digital Rights Management (DRM)
- State-of-the-art:
  - Stanley Chow, Philip A. Eisen, Harold Johnson, and Paul C. van Oorschot. A White-Box DES Implementation for DRM Applications. In Security and Privacy in Digital Rights Management, ACM CCS-9 Workshop, DRM 2002, volume 2696 of LNCS, pages 1–15. Springer, 2002.
  - Stanley Chow, Philip A. Eisen, Harold Johnson, and Paul C. van Oorschot. White-Box Cryptography and an AES Implementation. In Kaisa Nyberg and Howard M. Heys, editors, Selected Areas in Cryptography, volume 2595 of LNCS, pages 250–270, 2002.

Attacks on White-Box Cryptography

- Statistical attacks (similar to cryptanalysis techniques):
  - Louis Goubin, Jean-Michel Masereel, and Michaël Quisquater. Cryptanalysis of White Box DES Implementations. In Selected Areas in Cryptography, 14th International Workshop, SAC 2007, volume 4876 of LNCS, pages 278–295. Springer, 2007.
- Those which leverage techniques from grey-box analysis (i.e., side-channel or fault injection analyses), such as differential fault analysis, differential computation analysis, collision or mutual information, or high-order computational attacks
  - Joppe W. Bos, Charles Hubain, Wil Michiels, and Philippe Teuwen. Differential Computation Analysis: Hiding Your White-Box Designs is Not Enough. Cryptographic Hardware and Embedded Systems - CHES 2016 ,Santa Barbara, CA, USA, August 17-19, 2016, Proceedings, volume 9813 of LNCS, pages 215–236. Springer, 2016.
- Those which rely on Fourier transforms
  - Pascal Sasdrich, Amir Moradi, and Tim Güneysu. White-Box Cryptography in the Gray Box – A Hardware Implementation and its Side Channels. FSE 2016, Bochum, Germany, March 20-23, 2016, volume 9783 of LNCS, pages 185–203. Springer, 2016.

White-Boxing with Diffused-Input-Blocked-Output

Diffused-Input-Blocked-Output (DIBO)

- Function to white-box: $x \mapsto T(x + k^\star)$.
  - 8 bit to 32 bit
- T-box

$$T(x) = \begin{pmatrix} 02 \\ 01 \\ 01 \\ 03 \end{pmatrix} S(x) = \begin{pmatrix} 02S(x) \\ S(x) \\ S(x) \\ 03S(x) \end{pmatrix}$$

- Hiding elements = random bijections:
  - Linear permutation
  - Blocked bijection

$$x \mapsto O_{k^\star}(x) = B \circ \phi \circ T(x + k^\star). \qquad (2)$$

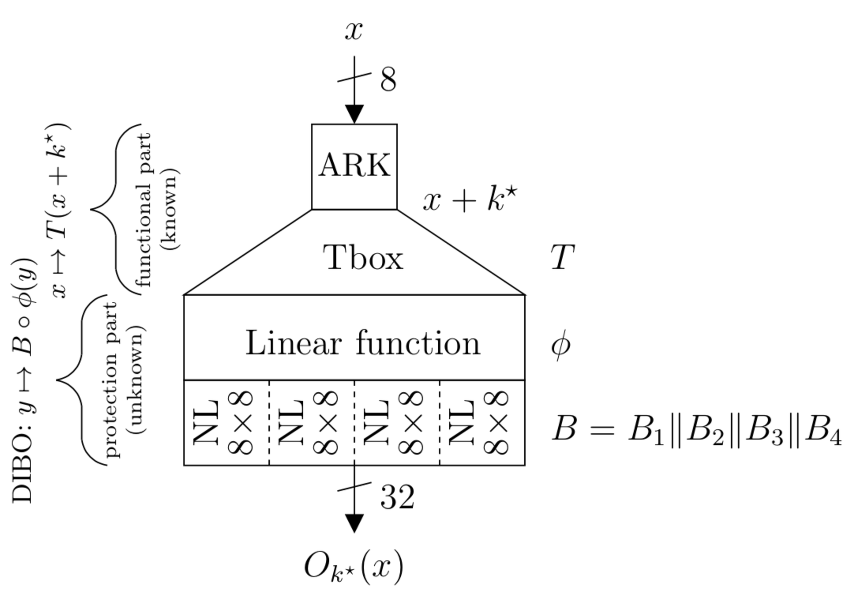


Figure 1: White-box protection $O_{k^\star}$ (equation (2)) of $x \in \mathbb{F}_2^8 \mapsto T(x + k^\star) \in \mathbb{F}_2^{32}$ (where T is known but $k^\star$ is one byte of the secret key), with DIBO function $B \circ \phi$ (i.e., the internal encoding). Notice that "NL" stands for the non-linear B_i, for $1 \leq i \leq 4$

See also:

ISO/IEC DTR 24485.3
Information technology — Security techniques — Security properties, test and evaluation guidance for white box cryptography

Our Distinguisher

Distinguisher: peeling the functional part

- Under key hypothesis
  - y is a 32-bit word:

$$\mathcal{A}_k : y \mapsto O_{k^\star}\left(T^{-1}(y) + k\right) = B \circ \phi \circ T\left(T^{-1}(y) + (k + k^\star)\right)$$

  - x is a 8-bit word:

$$A_k : x \mapsto \mathcal{A}_k(02x, x, x, 03x) = O_{k^\star}\left(S^{-1}(x) + k\right) = B \circ \phi \circ T\left(S^{-1}(x) + (k + k^\star)\right).$$

Distinguishers concept

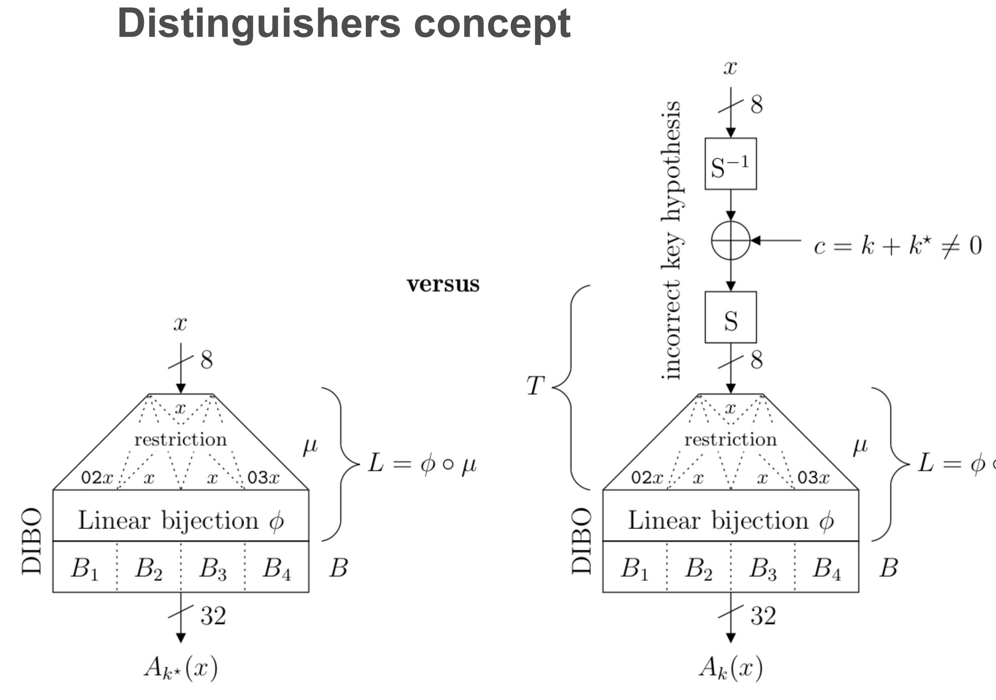


Figure 2: Two WBC situations to be distinguished, cases $A_{k^\star}$ and A_k, for $k \neq k^\star$.

Two distinguishers

- **Definition 4** (Spectral distinguisher of Sasdrich et al. [SMG16, §4.4 at page 200]).

$$\hat{k} = \underset{k \in \mathbb{F}_2^8}{\operatorname{argmin}} \sum_{u \in \mathbb{F}_2^8} \sum_{\substack{v \in \mathbb{F}_2^{32} \\ s.t.\ w_H(v)=1}} \left|W_{A_k}(u, v)\right|.$$

Grey-box context

- **Definition 5** (Our spectral distinguisher for WBC based on DIBO).

$$\hat{k} = \underset{k \in \mathbb{F}_2^8}{\operatorname{argmax}} \#\left\{W_{A_k}(u, v) = 0 \mid u \in \mathbb{F}_2^8, v \in E\right\}$$

WBC context

where:

$$E = \{(\mathbb{F}_2^8, 0, 0, 0), (0, \mathbb{F}_2^8, 0, 0), (0, 0, \mathbb{F}_2^8, 0), (0, 0, 0, \mathbb{F}_2^8)\} \subset \mathbb{F}_2^{32},$$

considering that $(\mathbb{F}_{2^8}, 0, 0, 0)$ stands for $\mathbb{F}_{2^8} \times \{0\}^3$ where 0 is the zero in $\mathbb{F}_{2^8}$.

recall the Walsh transform: $W_F(u, v) = \sum_{x \in \mathbb{F}_2^n} (-1)^{v \cdot F(x) + u \cdot x}$

[SMG16] Pascal Sasdrich, Amir Moradi, and Tim Güneysu. White-Box Cryptography in the Gray Box – A Hardware Implementation and its Side Channels. FSE 2016, Bochum, Germany, March 20-23, 2016, volume 9783 of LNCS.

Comparison between the two distinguishers

- Our new distinguisher is motivated:

Lemma 2. *Let $1 \leq i \leq 4$, and let $F = B_i \circ L_i$ a function from $\mathbb{F}_2^8$ to $\mathbb{F}_2^8$ (corresponding to the ith output of $A_{k^\star}$). Then the number of zeros in $W_F(u, v)$ is at least $2^8 - 2^{\mathrm{rank}(L_i)}$.*

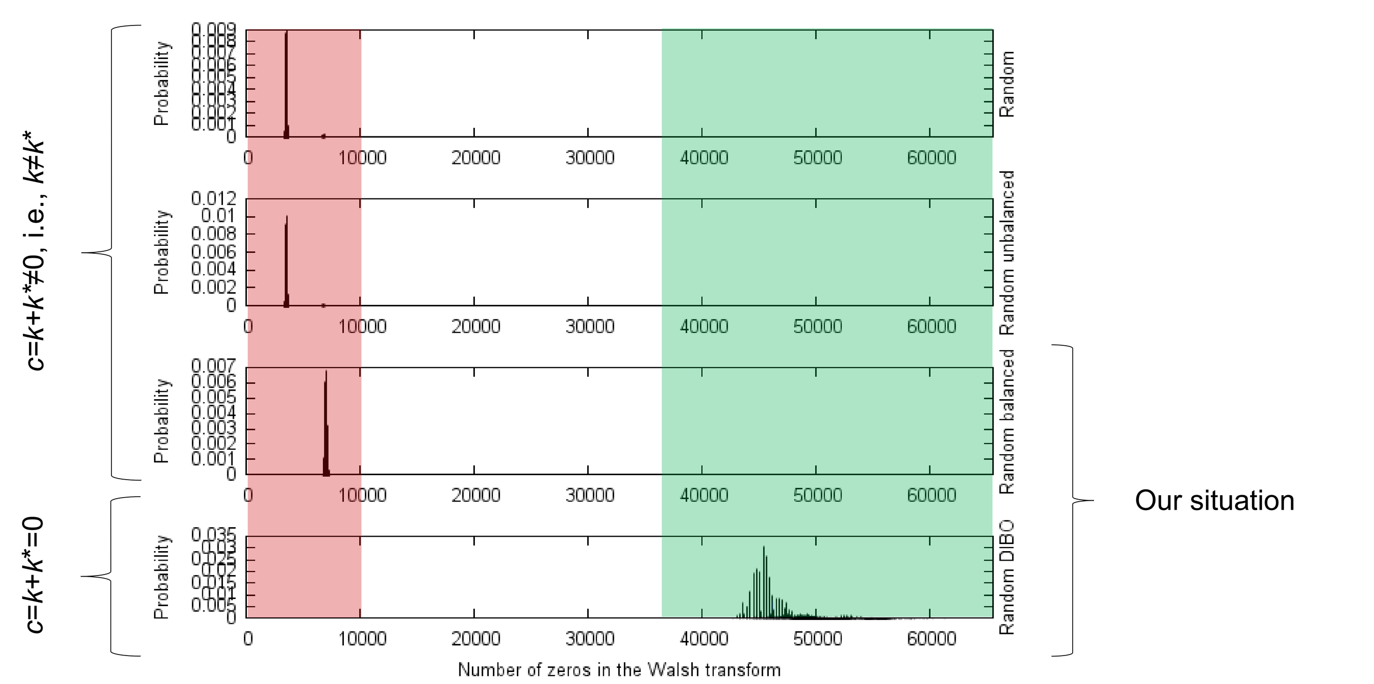
- Our new distinguisher is more tractable: Computational complexity = 2^{29} vs 2^{51}
- We provide a proof that our distinguisher always works when at least one L_i is neither bijective nor null
  - This hints for the countermeasure
  - And at the same time proves its theoretical fundation
- Notice that the fact our distinguisher works implies Sasdrich et al.'s working principle, by a Cauchy-Schwarz argument

$$\sum_{x \in \mathbb{F}_{2^8}} \sum_{i=1}^{32} \left|W_{(A_k)_i}(x)\right| \leq \sqrt{32 \times (2^8)^2 \times N}$$

where N equals the number of non-zeros among the values of the Walsh transform of the coordinate functions.

Our Attack & our Countermeasure

Intuition (simulations) $\#\left\{W_{A_k}(u, v) = 0 \mid u \in \mathbb{F}_2^8, v \in E\right\}$



Mathematical proof

Let r be the rank of L_i, $0 \leq r < 8 = n$

- Let g_c such that: $W_{g_c}(u) = \sum_{x \in \mathbb{F}_{2^n}} (-1)^{tr(ux) + g(tr(b_1(x^{-1}+c)^{-1}), \ldots, tr(b_r(x^{-1}+c)^{-1}))}$
- When c=k+$k^\star$=0:

When $c = k + k^\star = 0$, we can state a simple lower bound on the number of 0 of W_{g_0} depending only on r. According to Lemma 2, let g be an r-variable Boolean function. The size of $(W_{g_0})^{-1}(0)$ is larger than or equal to $2^n - 2^r$.

Remark 6. This result actually works for any value $0 \leq r \leq 8$.

Therefore, we need now to prove that the number of zeros in W_{g_c} is strictly less than $2^n - 2^r$ when $c \neq 0$.

- When c≠0, i.e., k≠$k^\star$:

Apply **Theorem 1** to the case f=g_c where c≠0

Theorem 1 ([BC99]). *Let f be a Boolean function over $\mathbb{F}_{2^n}$. The size of the Fourier-Hadamard support $\{u \in \mathbb{F}_{2^n}; \hat{f}(u) = \sum_{x \in \mathbb{F}_{2^n}} f(x)(-1)^{tr(ux)} \neq 0\}$ is larger than or equal to $2^{d^\circ_{alg} f}$.*

[BC99] Anna Bernasconi and Bruno Codenotti. Spectral Analysis of Boolean Functions as a Graph Eigenvalue Problem. IEEE Trans. Computers, 48(3):345–351, 1999.

Illustration of our mathematical proof

Soundness of the distinguisher:

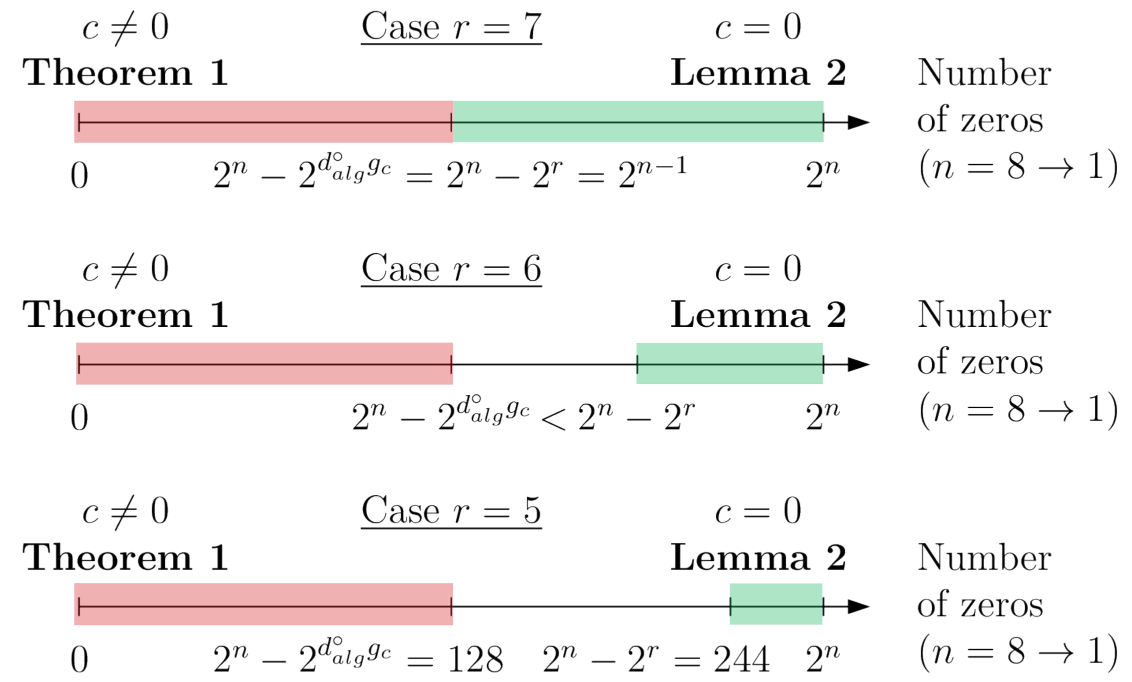
Wrong keys and correct key areas do not overlap.

Nota bene:

In our proof, the number of zeros for c≠0 is actually a probability.

It is = 2^{n-1} with proba $1-2^{-n}$ when r<7.

Cf. our **Theorem 2**



... even better distinguishing conditions when r is smaller

Countermeasure

5.1 Average insecurity of DIBO on AES

From the previous analysis, one can state the following

Countermeasure 1. *A DIBO obfuscation scheme is* ***immune*** *to our attack provided* ***all four*** *linear functions L_i, $1 \leq i \leq 4$, are invertible.*

Indeed, in such conditions, the use of Lemma 2 is no longer relevant.

In general, many linear $L_i : \mathbb{F}_2^8 \rightarrow \mathbb{F}_2^8$ are permutations. Namely, the number of permutations is $\prod_{i=0}^{7}(2^8 - 2^i)$, therefore the proportion of invertible linear mappings in $\mathbb{F}_2^8$ is $2^{-8^2} \prod_{i=0}^{7}(2^8 - 2^i) \approx 0.290$.

But now, for a DIBO obfuscations scheme to be attackable by our distinguisher, it suffices that at least one L_i is non-invertible. Hence the proportion of vulnerable DIBO is:

$$1 - \left(\prod_{i=0}^{7}\left(1 - 2^{i-8}\right)\right)^4 \approx 0.993. \qquad (12)$$