Analysis and Improvement of Differential Computation Attacks against Internally-Encoded White-Box Implementations

Matthieu Rivain ¹  Junwei Wang ¹,²,³

¹CryptoExperts  ²University of Luxembourg  ³University Paris 8

CHES 2019, Atalanta
White-Box Threat Model

- **Goal:** to extract a cryptographic key, ···
- **Where:** from a software impl. of cipher
- **Who:** malwares, co-hosted applications, user themselves, ···
- **How:** *(by all kinds of means)*
  - analyze the code
  - spy on the memory
  - interfere the execution
  - ···

In theory: no provably secure white-box scheme for standard block ciphers.
White-Box Threat Model

- **Goal:** to extract a cryptographic key, · · ·
- **Where:** from a software impl. of cipher
- **Who:** malwares, co-hosted applications, user themselves, · · ·
- **How:** (by all kinds of means)
  - analyze the code
  - spy on the memory
  - interfere the execution
  - · · ·

*In theory:* no provably secure white-box scheme for standard block ciphers.
Typical Applications

Digital Content Distribution
videos, music, games, e-books, ···

Host Card Emulation
mobile payment without a secure element

In practice:
heuristic solutions / security through obscurity
Typical Applications

**Digital Content Distribution**

videos, music, games, e-books, ...

**Host Card Emulation**

mobile payment without a *secure element*

In practice: heuristic solutions / security through obscurity
Internal Encoding Countermeasure [SAC02]

1. Represent the cipher into a network of transformations
**Internal Encoding Countermeasure** [SAC02]

1. Represent the cipher into a network of transformations
2. Obfuscate the network by encoding adjacent transformations
Internal Encoding Countermeasure [SAC02]

1. Represent the cipher into a network of transformations
2. Obfuscate the network by encoding adjacent transformations
3. Store the encoded transformations into look-up tables
Attacks in This Talk

1. Differential Computation Analysis
2. Collision Attack
Differential Computation Analysis [CHES16]

gray-box model
side-channel leakages (*noisy*)
e.g. power/EM/time/···

white-box model
computational leakage (*perfect*)
e.g. registers/accessed memory/···
Differential Computation Analysis [CHES16]

Differential power analysis techniques on computational leakages

collect traces → group by predictions → average trace → differential trace

$\varphi_k() = 0$

$\varphi_k() = 1$

Implying strong linear correlation between the sensitive variables and the leaked samples in the computational traces.
DCA Attack Limitations

1. The seminal work [CHES16] lacks in-depth understanding of DCA

2. The follow-up analysis [ACNS18] is
   - *partly* experimental (in particular for wrong key guesses)
   - *Only* known to work on nibble encodings
   - *Only* known to work on the first and last rounds
   - Success probability is unknown

3. The computational traces are only sub-optimally exploited
A key-dependent \((n, m)\) selection function \(\varphi_k\) in a block cipher

A random selected \(m\)-bit bijection \(\varepsilon\)

\(\varepsilon \circ \varphi_k\), as a result of some table look-ups, is leaked in the memory

To exploit the leakage of \(\varepsilon \circ \varphi_k\), it is necessary that \(n > m\)
DCA Analysis

Based on well-established theory – *Boolean correlation*, instead of *difference of means*: for any key guess $k$

$$\rho_k = \text{Cor}(\quad , \quad )$$

DCA success (roughly) requires:

$$|\rho_k^*| > \max_k \times |\rho_k| \times 10$$
DCA Analysis

Based on well-established theory – *Boolean correlation*, instead of *difference of means*: for any key guess $k$

$$\rho_k = \text{Cor}(\varphi_k(\cdot)[i], \varphi_k(\cdot)[j])$$

DCA success (roughly) requires:

$$|\rho_k^*| > \max_k |\rho_k| \times 10$$
DCA Analysis

Based on well-established theory – *Boolean correlation*, instead of *difference of means*: for any key guess \( k \)

\[
\rho_k = \text{Cor}(\varphi_k(\cdot)[i], \varepsilon \circ \varphi_k^*(\cdot)[j])
\]

DCA success (roughly) requires:

\[
|\rho_k^*| > \max_k \times |\rho_k| \times 10
\]
DCA Analysis

Based on well-established theory – *Boolean correlation*, instead of *difference of means*: for any key guess $k$

$$\rho_k = \text{Cor}( \varphi_k(\cdot)[i], \varepsilon \circ \varphi_k^*(\cdot)[j] )$$

DCA success (roughly) requires:

$$|\rho_k^*| > \max_{k^*} |\rho_{k^*}|$$
\( \rho_{k^*} \) and \( \rho_{k^\times} \): Distributions

- **Ideal** assumption: \( (\varphi_k)_k \) are mutually independent random \((n, m)\) functions
\( \rho_{k^*} \) and \( \rho_{k^\times} \): Distributions

- **Ideal** assumption: \( (\varphi_k)_k \) are mutually independent random \((n,m)\) functions

Correct key guess \( k^* \),

\[
\rho_{k^*} = 2^{2^{-m}N^* - 1}
\]

where

\[
N^* \sim \mathcal{HG}(2^m, 2^{m-1}, 2^{m-1})
\]

*Only* depends on \( m \).
\( \rho_{k^*} \) and \( \rho_{k^\times} \): Distributions

- **Ideal** assumption: \((\varphi_k)_k\) are mutually independent random \((n, m)\) functions

Correct key guess \( k^* \),

\[
\rho_{k^*} = 2^{2^m} N^* - 1
\]

where

\[
N^* \sim \mathcal{H}G(2^m, 2^{m-1}, 2^{m-1}) .
\]

*Only depends on \( m \).*

Incorrect key guess \( k^\times \),

\[
\rho_{k^\times} = 2^{2^n} N^\times - 1
\]

where

\[
N^\times \sim \mathcal{H}G(2^n, 2^{n-1}, 2^{n-1}) .
\]

*Only depends on \( n \).*
Lemma

Let $B(n)$ be the set of balanced $n$-bit Boolean function. If $f \in B(n)$ and $g \xleftarrow{\$} B(n)$ independent of $f$, then the balanceness of $f + g$ is $B(f + g) = 4 \cdot N - 2^n$ where $N \sim \mathcal{HG}(2^n, 2^{n-1}, 2^{n-1})$ denotes the size of $\{x : f(x) = g(x) = 0\}$.

With

$$\text{Cor}(f + g) = \frac{1}{2^n} B(f + g)$$

$$\Rightarrow$$

$$\rho_{k^*} = 2^{2-m} N^* - 1 \quad \text{and} \quad \rho_{k^x} = 2^{2-n} N^x - 1$$

where $N^* \sim \mathcal{HG}(2^m, 2^{m-1}, 2^{m-1})$ and $N^x \sim \mathcal{HG}(2^n, 2^{n-1}, 2^{n-1})$.
$\rho_{k^*}$ and $\rho_{k^\times}$: Distributions

PMF

$\times$ $\rho_{k^*}$ modeled

$\times$ $\rho_{k^\times}$ modeled

$n = 8, m = 4$
$\rho_k^*$ and $\rho_k^\times$: Distributions

$n = 8, m = 4$

PMF

Counts
DCA Success Rate: $|\rho_{k^*}| > \max_k |\rho_k|$
DCA Success Rate: $|\rho_{k^*}| > \max_k |\rho_k|$
Attack a NSC Variant: a White-Box AES

- *Byte* encoding protected
- DCA has failed to break it *before this work*
Attack a NSC Variant: a White-Box AES

- *Byte* encoding protected
- DCA has failed to break it *before this work*
- Our approach: target a output byte of MixColumn in the first round

\[
\begin{align*}
\phi^k_1 | | k_2 (x_1 | | x_2) &= 2 \cdot \text{Sbox}(x_1 \oplus k_1) \\
&\oplus 3 \cdot \text{Sbox}(x_2 \oplus k_2)
\end{align*}
\]

\[
\epsilon' = \epsilon \circ \oplus c, n = 16, m = 8, |K| = 2^{16}.
\]
Attack a NSC Variant: a White-Box AES

- Byte encoding protected
- DCA has failed to break it before this work
- Our approach: target a output byte of MixColumn in the first round

\[
\begin{array}{c}
\chi_1 \\
\chi_2 \\
0 \\
0 \\
\end{array}
\quad \xrightarrow{\text{ARK,SB}}
\quad \begin{array}{c}
\chi_1 \\
\chi_2 \\
0 \\
0 \\
\end{array}
\]

\[
\begin{align*}
\text{Sbox}(x_1 \oplus k_1) & \quad \text{Sbox}(x_2 \oplus k_2) & \quad \text{Sbox}(k_3) & \quad \text{Sbox}(k_4)
\end{align*}
\]
Attack a NSC Variant: a White-Box AES

- *Byte* encoding protected
- DCA has failed to break it *before this work*
- Our approach: target a output byte of MixColumn in the first round

\[
\begin{align*}
X^1 & \rightarrow \text{ARK,SB} & \rightarrow \text{SR} \quad \text{Sbox}(x_1 \oplus k_1) & \quad \text{Sbox}(x_2 \oplus k_2) & \quad \text{Sbox}(k_3) & \quad \text{Sbox}(k_4)
\end{align*}
\]
Attack a NSC Variant: a White-Box AES

- Byte encoding protected
- DCA has failed to break it before this work
- Our approach: target a output byte of MixColumn in the first round

\[ 2 \cdot \text{Sbox}(x_1 \oplus k_1) \oplus 3 \cdot \text{Sbox}(x_2 \oplus k_2) \oplus \text{Sbox}(k_3) \oplus \text{Sbox}(k_4) \]
Attack a NSC Variant: a White-Box AES

- Byte encoding protected
- DCA has failed to break it before this work
- Our approach: target a output byte of MixColumn in the first round

\[ \phi_{k_1} = \phi \oplus c, \quad n = 16, \quad m = 8, \quad |K| = 2^{16}. \]
Attack a NSC Variant: a White-Box AES

- **Byte** encoding protected
- DCA has failed to break it *before this work*
- Our approach: target a output byte of MixColumn in the first round

\[
\varphi_{k_1||k_2}(x_1||x_2) = 2 \cdot \text{Sbox}(x_1 \oplus k_1) \oplus 3 \cdot \text{Sbox}(x_2 \oplus k_2)
\]

\[
\varepsilon' = \varepsilon \circ \oplus c ,
\]

\[
n = 16, m = 8, |\mathcal{K}| = 2^{16}.
\]
Attack a NSC Variant: a White-Box AES

- Attack results: \( \sim 1800 \) traces

- Similar attack can be applied to a “masked” white-box implementation, which intends to resist DCA.
Attacks in This Talk

1. Differential Computation Analysis

2. Collision Attack
Collision Attack

$N$ inputs & raw traces
Collision Attack

\( N \) inputs & raw traces \( \binom{N}{2} \) collision predictions & traces

\[ \psi_k(x_1, x_2) \]

\[ \psi_k(x_1, x_3) \]

\[ \psi_k(x_1, x_4) \]

\[ \psi_k(x_2, x_3) \]

\[ \psi_k(x_2, x_4) \]

\[ \psi_k(x_3, x_4) \]

\[ \psi_k(x_1, x_2) := (\varphi_k(x_1) = \varphi_k(x_2)) \]
Collision Attack

$N$ inputs & raw traces \quad \binom{N}{2}$ collision predictions & traces

\[
\psi_k(x_1, x_2) := \left( \varphi_k(x_1) = \varphi_k(x_2) \right)
\]
Collision Attack: Explanation

Based on the principle:

\[ \varphi_k(x_1) = \varphi_k(x_2) \iff \varepsilon \circ \varphi_k(x_1) = \varepsilon \circ \varphi_k(x_2) \]

Trace Complexity:

\[ N = O\left(2^\frac{m}{2}\right) \]
Collision Attack: Explanation

Predictions

key guesses

\[ k_1 \]

\[ \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
\end{array} \]

\[ \begin{array}{cccccc}
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\end{array} \]

\[ k^* \text{ "collides" } \land \forall k^\times, k^* \text{ and } k^\times \text{ are not "isomorphic"} \]

\[ \Rightarrow N = O \left( 2^\frac{m}{2} \right) \]
Collision Attack: Explanation

$k^* \text{ "collides" } \land \forall k^\times, k^*$ and $k^\times$ are not "isomorphic"

$\Rightarrow N = O\left(2^{\frac{m}{2}}\right)$
Attack the NSC Variant

- Same to DCA: targeting at one 1-st round MixColumn output byte
- Attack results: 60 traces
Conclusion

- DCA against internal encodings has been analysed in depth
  - Allows to attack wider encodings
- Computation traces have been further exploited
  - Showcase to attack variables beyond the first round of the cipher
  - New class of collision attack with very low trace complexity
- Hence, protecting AES with internal encodings in the beginning rounds is insufficient
Thank You!

ia.cr/2019/076