SCIS 2022 2022 Symposium on Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 – 21, 2022
The Institute of Electronics, Information and Communication Engineers

ハードウェア実装 AES に対する Aggregated Mono-Bit model モデルを利用した深層学習サイドチャネル攻撃

Deep Learning Side-Channel Attack using Aggregated Mono-Bit model against Hardware-Implemented AES

橋本 尚志 * 福田 悠太 * 吉田 康太 * 黒田 訓宏 *
Hisashi Hashimoto Yuta Fukuda Kota Yoshida Kunihiro kuroda

藤野 毅 *
Takesi Fujino

キーワード サイドチャネル攻撃，ハードウェア実装 AES，深層学習

あらまし

ループ型アーキテクチャを用いたハードウェア実装 AES に対するプロファイリング深層学習サイドチャネル攻撃 (DL-SCA) を評価する．DL-SCA は AES 回路の内部レジスタの状態遷移が消費電力に影響を与えることを利用し，電力波形から内部レジスタの遷移を深層ニューラルネットワーク (DNN) を用いて予測することで攻撃を行う．レジスタの遷移の XOR 値 (HD-ID) を DNN に予測させると，一部の HD-ID 値が学習できない (ラベル欠損問題が発生する) ため 0,4,8,12Byte 目の部分鍵に対して攻撃できないことが報告されている．

本稿で評価する Aggregated mono-bit モデルでは，プロファイリング時に HD-ID をバイナリ表現した各 bit ごとにその値を予測する Mono-bit DNN モデル (f^n, $n \in \{1, 2, \ldots, 8\}$) を訓練する．攻撃時には図 1 のように消費電力波形 $t \in T$ が与えられた時の全ての DNN モデルの出力確率 $p(t|k_c^n)$ を式 (1) により集約することで，ベイズ推定を行う際の条件付き確率が求まる．攻撃波形セット T が与えられたときに秘密鍵が k_c である確率 $p(k_c|T)$ を全ての候補鍵から比較することで秘密鍵を明らかにする．

$$p(t|k_c) = p(t|k_c^1) \times p(t|k_c^2) \times \cdots \times p(t|k_c^8) \qquad (1)$$

未対策 AES に対して攻撃を行った結果を図 2 に示す．プロファイリングデバイスの鍵が 1 種類 (1-key) では，従来の HD-ID モデルが 12byte，Aggregated mono-bit モデルでも 13byte の導出にとどまった．プロファイリングデバイスの鍵を 3 種類に増やす (3-key) と，従来の HD-ID モデルでは変わらず 12byte であったが Aggregated mono-bit モデルでは 16byte 全ての鍵を明らかにすることができた．

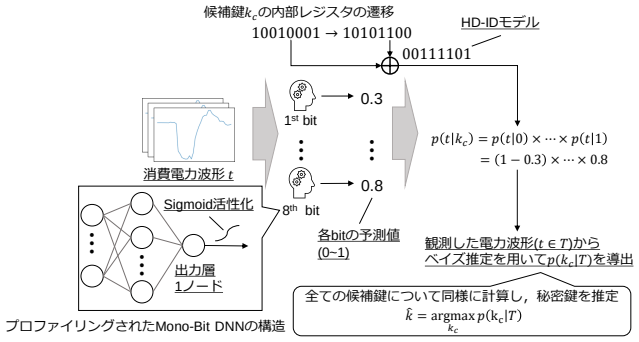

図 1: Aggregated mono-bit モデルを用いたプロファイリング DL-SCA

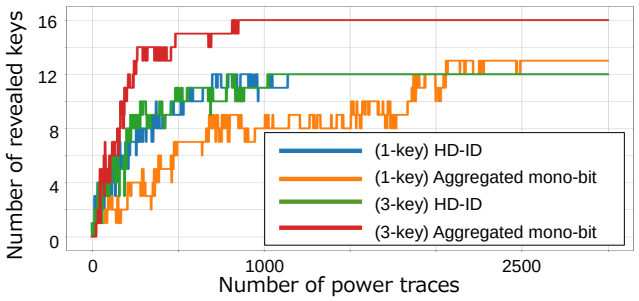

図 2: 未対策 AES に対する攻撃結果

* 立命館大学, 〒 525-8577 滋賀県草津市野路東 1-1-1, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga, Japan