A Novel Use of Kernel Discriminant Analysis as a Higher-Order Side-Channel Distinguisher

Xinping Zhou\textsuperscript{1}, Carolyn Whitnall\textsuperscript{2}, Elisabeth Oswald\textsuperscript{2}, Degang Sun\textsuperscript{1} and Zhu Wang\textsuperscript{1}

\textsuperscript{1}Chinese Academy of Sciences
\textsuperscript{2}University of Bristol

CARDIS 2017, Lugano, Switzerland
14\textsuperscript{th} November 2017
Outline

1. Introduction
2. Preliminary
3. Methodology
4. Discussion
5. Conclusions and Future Perspectives
Introduction

KDA as a Higher-Order Side-Channel Distinguisher
Introduction

$L(k, x; r)$

A \xrightarrow{x} y \xrightarrow{O} k

KDA as a Higher-Order Side-Channel Distinguisher
Differential side channel

Secret key $k^*$

Device

$L(F_{k^*}(X)) + \text{noise}$

Known input $X$

Model

$M(F_k(X))$

Hypothetical Key $k$

$D \rightarrow \text{Values}$
Introduction

Masking countermeasure

Device

Model

Known input $X$

Secret key $k^* \oplus r$

Hypothetical Key $k$

$D$

$L(F_{k^*}(X)) + noise$

$M(F_k(X))$

Values

KDA as a Higher-Order Side-Channel Distinguisher

Slide 6 of 27
### Logic motivation of this work

- Linear Discriminant Analysis (LDA) was used for first-order dimensionality reduction. (@ CHES 2008 by Standaert et al.)
- LDA was used as first order distinguisher. (@ RFIDsec 2016 by Mahmudlu et al.)
- Kernel Discriminant Analysis (KDA) was successfully used for dimensionality reduction (or POI selection) in higher-order implementation. (@ CARDIS 2016 by Cagli et al.)
- KDA is proposed as higher-order distinguisher in this work.

### References

Masking countermeasure (boolean masking)

- Sensitive value is to split into several shares
  \[ s = r_0 \otimes r_1 \otimes ... \otimes r_d \]

- The whole leakages are \( l = (l_0, l_1, ..., l_d) \) with
  \[ l_0 = L_0 \circ (s \oplus r_1 \oplus ... \oplus r_d) + \varepsilon_0 \]
  \[ l_i = L_i \circ (r_i) + \varepsilon_i, \quad \text{for } 1 \leq i \leq d. \]
Higher-order DPA

\[ R^{(d+1)\ell} \xrightarrow{CF} R^{\ell^{d+1}} \xrightarrow{D} k^* \]
Preliminary

Linear discriminant analysis

- LDA seeks the directions on that the labeled data have max ratio of between-cluster scatter and within-class scatter.
  - LDA is used as reduction tool in profiled-analysis in SCA.
  - Based on the ratio of between-cluster scatter and within-class scatter, it can distinguish the correct key hypothesis and wrong ones.
Linear discriminant analysis with kernels

$R^U \xrightarrow{\Phi} \mathcal{F} \xrightarrow{LDA} R^C$

$KDA$
Kernel discriminant analysis

- KDA seeks the optimal directions in a non-linear space.
  - KDA is used as dimensionality reduction tool in higher-order profiled-analysis in SCA.
  - The eigenvectors with largest eigenvalues are selected in the dimensionality reduction.
Methodology

Natural common ground

\[ R^{(d+1)} \ell \xrightarrow{CF} R^{\ell^{d+1}} \xrightarrow{D} k^* \]

\[ R^U \xrightarrow{\Phi} F \xrightarrow{LDA} R^C \]

KDA as a Higher-Order Side-Channel Distinguisher
Basic idea of KDA distinguisher

- If key hypothesis is correct, the partition of the whole traces based on the intermediate value corresponds with the real partition.
- In this case, it is easy to find the max ratio of between-cluster distance and inner-cluster distance.
- Otherwise, the clusters are difficult to separate.
Detailed procedure of KDA distinguisher

- For each key hypothesis $k \in \mathcal{K}$, do the following:
  - Calculate the intermediate value $z_i = F_k(x_i)$ for each plaintext.
  - Map $z_i$ to a power model prediction $m_i$, given by $M(z_i)$.
  - Compute the between-class scatter matrix $\mathbf{M}$ and the within-class scatter matrix $\mathbf{N}$, and regularize $\mathbf{N}$ by $\mathbf{N} = \mathbf{N} + \mu \mathbf{I}$.
  - Eigen-decompose the matrix $\mathbf{N}^{-1} \mathbf{M}$. Return the largest eigenvalue as the distinguisher score $D_k$ for $k$.
- Rank the pairs $(k, D_k)$ according to $D_k$.
- Output the key hypothesis $k$ with the largest $D_k$ as the best guess on the true subkey.
Methodology

Theoretical Rationale

- The effectiveness of the implicit projection.
- The effectiveness of LDA as a distinguisher in the first-order scenario.
Methodology

Experimental Validation

- Real traces from DPA contest v4 (for second-order analysis).
  - Attack target: XOR result of masked S-box output and masked value of next sub-plaintext in RSM scheme.
- Simulated multivariate leakages (for second-order and third-order analysis).
  - Attack target: XOR result of random shares.
- Kernel function (might not be optimal)
  - The kernel function is $K(x, y) = (x \cdot y)^{d+1}$.
  - Regularization factor $\mu = 100,000$
Methodology

KDA on second-order simulated masked implementation with $\sigma = 1$.

Second-order KDA distinguisher key recovery

- Correct key candidate
- Wrong key candidates

KDA distinguisher score vs. Attack sample
Methodology

KDA on second-order simulated masked implementation with $\sigma = 1$. 

![Graph showing the comparison between Second-order KDA and Second-order KDA (LSB model). The x-axis represents the attack sample, and the y-axis represents the guessing entropy. The graph illustrates the performance of both methods over the attack sample.]
Methodology

Second-order with KDA on DPA v4

The graph shows the Guessing entropy against the Attack sample. The red line represents the Second-order KDA, indicating a significant reduction in entropy as the number of attack samples increases.
Methodology

Third-order with KDA on simulated masked implementation with $\sigma = 0.01$. 

![Graph showing guessing entropy against attack sample numbers with a line labeled Third-order KDA.](image)
Discussions

### Computation Complexity

- **Time Complexity:**
  - Classical higher-order DPA: $O(N^{d+1})$.
  - KDA method: $O(N^2(N + (d + 1)\ell))$.

- **Space Complexity:**
  - Classical higher-order DPA: $N^{d+1}$.
  - KDA method: $2N^2$. 
Discussions

Power Model

- Classical higher-order DPA: Standard proportional power models.
- KDA method: Flexible clustering power models.
Limitations and Possibilities

- Classical higher-order DPA using the ‘normalised product’ combining function with Hamming weight outperforms the KDA.
- It is interesting to deploy the KDA distinguisher in scenarios where higher order correlation DPA is likely to struggle.
Conclusions and Future Perspectives

Conclusions

- Extended KDA for application as distinguisher in masked implementation.
- Showed natural common ground between classical higher-order DPA and KDA.
- Reasoned about the soundness of a KDA-based distinguisher from theoretical perspective and empirically.
- Analyzed the substantial advantages of KDA over higher-order DPA on complexity and power model.
Conclusions and Future Perspectives

Future Perspectives

- Optimizing the parameters such as regularization factor.
- Exploring other kernel functions besides the polynomial function.
- Combining clustering power model in CHES 2015 proposed by Whitnall et al.

- Whitnall, C., Oswald, E. Robust profiling for DPA-style attacks. CHES 2015. 3-21.
Thank you for listening!

Full version available at https://eprint.iacr.org/2017/1051