Optimal Attacks for Multivariate and Multimodel Side-Channel Leakages

Nicolas Bruneau, Sylvain Guilley, Annelie Heuser,
Damien Marion and Olivier Rioul

Saturday August 20, 2016

PROOFS, UCSB, Santa Barbara

Outline

Presentation Outline

Facts

Side-channel leakages are:

- **multi-variate** . . . . . (in time)
- **multi-model** . . . . . (e.g., each bit leaks $\neq$)

Matrix Notations

- Q ..... number of queries,
- D ..... number of samples,
- S ..... number of models.

In matrix notation:

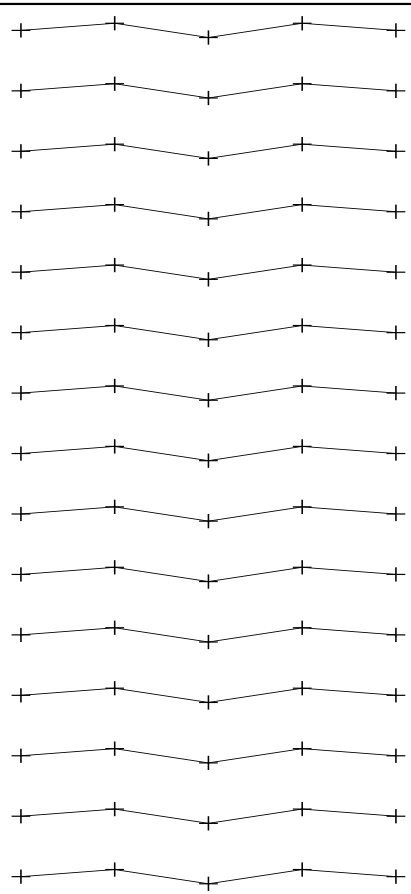
$$\mathbf{X} = \alpha \mathbf{Y}^{\star} + \mathbf{N} \tag{1}$$

where

- $\mathbf{X}$ is a matrix of size ..... $D \times Q$,
- α is a matrix of size ..... $D \times S$,
- $\mathbf{Y}^{\star}$ (the star means: "for the correct key $k = k^{\star}$") is a matrix of size ..... $S \times Q$,
- $\mathbf{N}$ is a matrix of size ..... $D \times Q$.

Examples of **X**

It is a matrix

Plaintext	Trace, **X**
0xe3e70682c2094cac629f6fbed82c07cd	
0x82e2e662f728b4fa42485e3a0a5d2f34	
0xd4713d60c8a70639eb1167b367a9c378	
0x23a7711a8133287637ebdcd9e87a1613	
0xe6f4590b9a164106cf6a659eb4862b21	
0x85776e9add84f39e71545a137a1d5006	
0xd71037d1b83e90ec17e0aa3c03983ca8	
0xf7b0b7d2cda8056c3d15eef738c1962e	
0x1759edc372ae22448b0163c1cd9d2b7d	
0x8c25166a1ff39849b4e1357d4a84eb03	
0x966e12778c1745a79a6a5f92cca74147	
0xcc45782198a6416d1775336d71eacd05	
0x4a5308cc3dfabc08935ddd725129fb7c	
0x79fdef7c42930b33a81ad477fb3675b8	
0xd7ab792809e469e6ec62b2c82648ee38	

Examples of **X**

It is a matrix

Plaintext	Trace, **X**				
0xe3e70682c2094cac629f6fbed82c07cd	8	9	5	3	7
0x82e2e662f728b4fa42485e3a0a5d2f34	2	8	8	8	5
0xd4713d60c8a70639eb1167b367a9c378	9	5	4	6	9
0x23a7711a8133287637ebdcd9e87a1613	9	7	0	6	4
0xe6f4590b9a164106cf6a659eb4862b21	6	8	2	7	1
0x85776e9add84f39e71545a137a1d5006	2	7	3	8	1
0xd71037d1b83e90ec17e0aa3c03983ca8	1	6	0	5	9
0xf7b0b7d2cda8056c3d15eef738c1962e	5	6	0	6	6
0x1759edc372ae22448b0163c1cd9d2b7d	5	3	3	9	0
0x8c25166a1ff39849b4e1357d4a84eb03	0	9	1	1	2
0x966e12778c1745a79a6a5f92cca74147	8	9	0	4	1
0xcc45782198a6416d1775336d71eacd05	2	2	6	3	1
0x4a5308cc3dfabc08935ddd725129fb7c	5	0	1	9	1
0x79fdef7c42930b33a81ad477fb3675b8	3	7	8	9	1
0xd7ab792809e469e6ec62b2c82648ee38	6	9	0	6	8

Examples of $\mathbf{Y}_k$

It is a matrix

Plaintext	1st byte	Bits of Sbox #0 ($\mathbf{Y}_k$ for $k = 0x00$)
0xe3e70682c2094cac629f6fbed82c07cd	0xbd	10111101
0x82e2e662f728b4fa42485e3a0a5d2f34	0x18	00011000
0xd4713d60c8a70639eb1167b367a9c378	0xbc	10111100
0x23a7711a8133287637ebdcd9e87a1613	0x7d	01111101
0xe6f4590b9a164106cf6a659eb4862b21	0xfd	11111101
0x85776e9add84f39e71545a137a1d5006	0x6f	01101111
0xd71037d1b83e90ec17e0aa3c03983ca8	0xc2	11000010
0xf7b0b7d2cda8056c3d15eef738c1962e	0x31	00110001
0x1759edc372ae22448b0163c1cd9d2b7d	0xff	11111111
0x8c25166a1ff39849b4e1357d4a84eb03	0x7b	01111011
0x966e12778c1745a79a6a5f92cca74147	0xa0	10100000
0xcc45782198a6416d1775336d71eacd05	0x6b	01101011
0x4a5308cc3dfabc08935ddd725129fb7c	0x10	00010000
0x79fdef7c42930b33a81ad477fb3675b8	0x6c	01101100
0xd7ab792809e469e6ec62b2c82648ee38	0x07	00000111

Examples of $\mathbf{Y}_k$

It is a matrix

Plaintext	1st byte	Bits of Sbox #0 ($\mathbf{Y}_k$ for $k = 0x01$)
0xe3e70682c2094cac629f6fbed82c07cd	0x4b	01001011
0x82e2e662f728b4fa42485e3a0a5d2f34	0x96	10010110
0xd4713d60c8a70639eb1167b367a9c378	0xb6	10110110
0x23a7711a8133287637ebdcd9e87a1613	0xc9	11001001
0xe6f4590b9a164106cf6a659eb4862b21	0xb7	10110111
0x85776e9add84f39e71545a137a1d5006	0xc5	11000101
0xd71037d1b83e90ec17e0aa3c03983ca8	0xd3	11010011
0xf7b0b7d2cda8056c3d15eef738c1962e	0x15	00010101
0x1759edc372ae22448b0163c1cd9d2b7d	0x10	00010000
0x8c25166a1ff39849b4e1357d4a84eb03	0x77	01110111
0x966e12778c1745a79a6a5f92cca74147	0x5a	01011010
0xcc45782198a6416d1775336d71eacd05	0xf2	11110010
0x4a5308cc3dfabc08935ddd725129fb7c	0xff	11111111
0x79fdef7c42930b33a81ad477fb3675b8	0x56	01010110
0xd7ab792809e469e6ec62b2c82648ee38	0x12	00010010

Examples of $\mathbf{Y}_k$

It is a matrix

Plaintext	1st byte	Bits of Sbox #0 ($\mathbf{Y}_k$ for $k = 0x02$)
0xe3e70682c2094cac629f6fbed82c07cd	0x8a	10001010
0x82e2e662f728b4fa42485e3a0a5d2f34	0x05	00000101
0xd4713d60c8a70639eb1167b367a9c378	0xda	11011010
0x23a7711a8133287637ebdcd9e87a1613	0x82	10000010
0xe6f4590b9a164106cf6a659eb4862b21	0x26	00100110
0x85776e9add84f39e71545a137a1d5006	0xf2	11110010
0xd71037d1b83e90ec17e0aa3c03983ca8	0xac	10101100
0xf7b0b7d2cda8056c3d15eef738c1962e	0x71	01110001
0x1759edc372ae22448b0163c1cd9d2b7d	0xd2	11010010
0x8c25166a1ff39849b4e1357d4a84eb03	0x7c	01111100
0x966e12778c1745a79a6a5f92cca74147	0x6e	01101110
0xcc45782198a6416d1775336d71eacd05	0xc5	11000101
0x4a5308cc3dfabc08935ddd725129fb7c	0xf3	11110011
0x79fdef7c42930b33a81ad477fb3675b8	0xf4	11110100
0xd7ab792809e469e6ec62b2c82648ee38	0x80	10000000

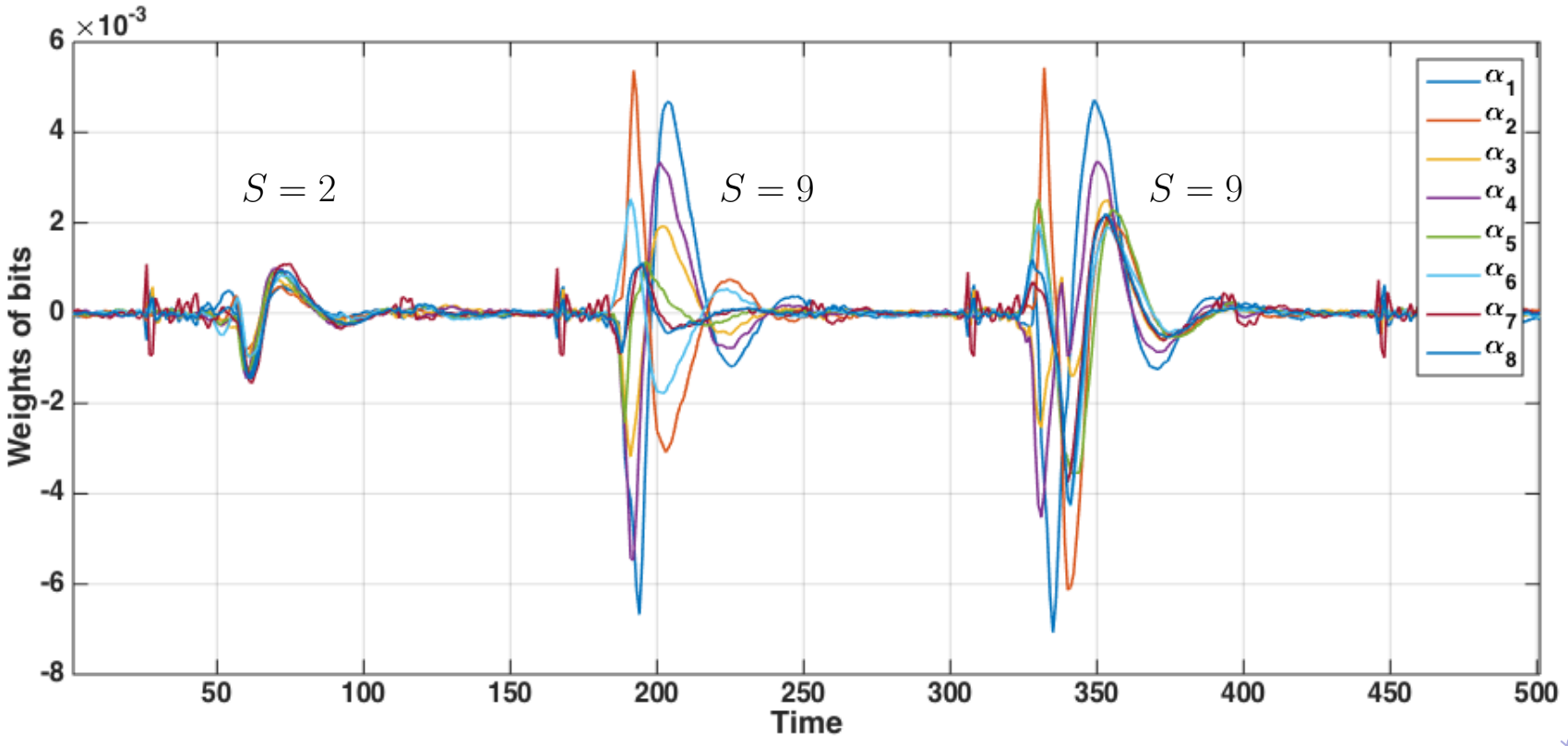
Examples of $\mathbf{Y}_k$

It is a matrix

Plaintext	1st byte	Bits of Sbox #0 ($\mathbf{Y}_k$ for $k = \texttt{0xff}$)
0xe3e70682c2094cac629f6fbed82c07cd	0x23	00100011
0x82e2e662f728b4fa42485e3a0a5d2f34	0x1f	00011111
0xd4713d60c8a70639eb1167b367a9c378	0x17	00010111
0x23a7711a8133287637ebdcd9e87a1613	0xce	11001110
0xe6f4590b9a164106cf6a659eb4862b21	0x1d	00011101
0x85776e9add84f39e71545a137a1d5006	0x99	10011001
0xd71037d1b83e90ec17e0aa3c03983ca8	0x5b	01011011
0xf7b0b7d2cda8056c3d15eef738c1962e	0x3e	00111110
0x1759edc372ae22448b0163c1cd9d2b7d	0x13	00010011
0x8c25166a1ff39849b4e1357d4a84eb03	0xb0	10110000
0x966e12778c1745a79a6a5f92cca74147	0x6c	01101100
0xcc45782198a6416d1775336d71eacd05	0x2d	00101101
0x4a5308cc3dfabc08935ddd725129fb7c	0xec	11101100
0x79fdef7c42930b33a81ad477fb3675b8	0xa0	10100000
0xd7ab792809e469e6ec62b2c82648ee38	0xc6	11000110

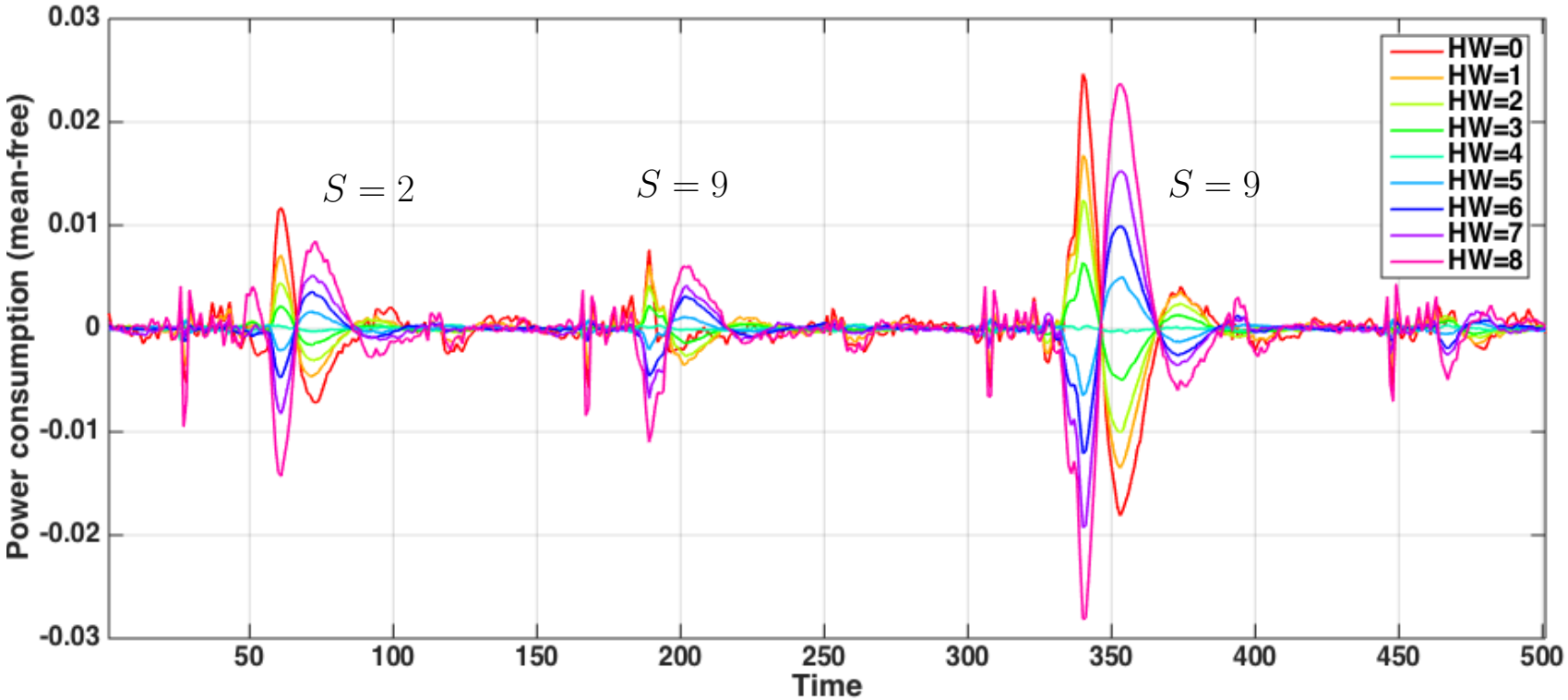
Real-World Example

The figure below shows power consumption traces taken from an ATMega smartcard—datasets are available from the DPA contest V4 team [TEL14] (knowing the mask).



Real-World Example

The figure below shows power consumption traces taken from an ATMega smartcard—datasets are available from the DPA contest V4 team [TEL14] (knowing the mask).



Question

What is the optimal distinguisher, when in Equation (1):

- α is known? ..... $\mathcal{D}_{\mathrm{ML}}(\mathbf{x}, \mathbf{t})$
- α is unknown? ..... $\mathcal{D}_{\mathrm{ML,sto}}(\mathbf{x}, \mathbf{t})$

Presentation Outline

Solution for α known I

Theorem

The optimal maximum likelihood (ML) distinguisher [HRG14] for Gaussian noise writes

$$\mathcal{D}_{\mathrm{ML}}(\mathbf{x}, \mathbf{t}) = \underset{k}{\operatorname{argmin}} \ \operatorname{tr}\Big((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\Big). \tag{2}$$

Notice that:

$$\operatorname{tr}\Bigg(\underbrace{(\mathbf{x} - \alpha\mathbf{y})^{\mathsf{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y})}_{Q\times Q \text{ matrix}}\Bigg) = \operatorname{tr}\Bigg(\underbrace{\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y})(\mathbf{x} - \alpha\mathbf{y})^{\mathsf{T}}}_{D\times D \text{ matrix}}\Bigg).$$

Proof.

From [HRG14] we have $\mathcal{D}_{\mathrm{ML}}(\mathbf{x}, \mathbf{t}) = \mathrm{argmax}_k \; p(\mathbf{x}|\mathbf{y}_k)$ where from (1) it is easily seen that $p(\mathbf{x}|\mathbf{y}_k) = p_{\mathbf{N}}(\mathbf{x} - \alpha\mathbf{y}_k)$. From the i.i.d. assumption the noise density $p_{\mathbf{N}}(\mathbf{n})$ is given by

$$p_{\mathbf{N}}(\mathbf{n}) = \prod_{q=1}^{Q} \frac{1}{\sqrt{(2\pi)^D |\det \Sigma|}} \exp -\frac{1}{2} {n_q}^{\mathsf{T}} \Sigma^{-1} n_q \tag{3}$$

$$= \frac{1}{(2\pi)^{DQ/2}} \frac{1}{(\det \Sigma)^{Q/2}} \exp -\frac{1}{2} \left(\sum_{q=1}^{Q} {n_q}^{\mathsf{T}} \Sigma^{-1} n_q \right) \tag{4}$$

$$= \frac{1}{(2\pi)^{DQ/2} (\det \Sigma)^{Q/2}} \exp -\frac{1}{2} \operatorname{tr}\left(\mathbf{n}^{\mathsf{T}} \Sigma^{-1} \mathbf{n} \right). \tag{5}$$

Thus $p_{\mathbf{N}}(\mathbf{x} - \alpha\mathbf{y}_k)$ is maximum when the expression $\operatorname{tr}\left(\mathbf{n}^{\mathsf{T}} \Sigma^{-1} \mathbf{n}\right)$ for $\mathbf{n} = \mathbf{x} - \alpha\mathbf{y}_k$ is minimum. □

Solution for α unknown

Theorem

The optimal stochastic multivariate attack is given by

$$\mathcal{D}_{\mathrm{ML,sto}}(\mathbf{x}, \mathbf{t}) = \underset{k \in \mathbb{F}_2^n}{\operatorname{argmax}} \ \operatorname{tr}\left(\mathbf{y}_k{}^{\mathsf{T}}(\mathbf{y}_k\mathbf{y}_k{}^{\mathsf{T}})^{-1}\mathbf{y}_k \ \ \mathbf{x}^{\mathsf{T}}\Sigma^{-1}\mathbf{x}\right). \tag{6}$$

for which the optimal value of α is given by

$$\alpha^{opt} = (\mathbf{x}\mathbf{y}_k{}^{\mathsf{T}})(\mathbf{y}_k\mathbf{y}_k{}^{\mathsf{T}})^{-1}. \tag{7}$$

Proof.

Let $\mathbf{x}' = \Sigma^{-1/2}\,\mathbf{x}$ and $\alpha' = \Sigma^{-1/2}\,\alpha$.
The optimal distinguisher minimizes the following expression over $\alpha \in \mathbb{R}^{D\times S}$:

$$\mathrm{tr}\left((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\right) = \mathrm{tr}\left((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1/2}\Sigma^{-1/2}(\mathbf{x} - \alpha\mathbf{y}_k)\right).$$

The minimization over α'_d yields $\alpha'_d = (\mathbf{x}'_d{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$ for all $d = 1, \ldots, D$. This gives $\alpha' = (\mathbf{x}'{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$ hence $\alpha = (\mathbf{x}{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$, which remarkably does *not* depend on Σ.
The minimized value of the distinguisher is thus

$$\begin{aligned}\min_{\alpha}\ \mathrm{tr}\left((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\right) &= \mathrm{tr}\left((\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)\right)\\ &= \mathrm{tr}\left((\mathrm{Id} - {\mathbf{y}_k}^{\mathrm{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1})^2\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right)\\ &= \mathrm{tr}\left(\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right) - \mathrm{tr}\left({\mathbf{y}_k}^{\mathrm{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}\ \mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right)\end{aligned}$$

where Id is the $D \times D$ identity matrix and $\mathrm{tr}\left(\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right)$ is a constant independent of k. □

Proof.

Let $\mathbf{x}' = \Sigma^{-1/2}\,\mathbf{x}$ and $\alpha' = \Sigma^{-1/2}\,\alpha$.
The optimal distinguisher minimizes the following expression over $\alpha \in \mathbb{R}^{D\times S}$:

$$\mathrm{tr}\Big((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\Big) = \mathrm{tr}\Big(\big(\Sigma^{-1/2}(\mathbf{x} - \alpha\mathbf{y}_k)\big)^{\mathrm{T}}\Sigma^{-1/2}(\mathbf{x} - \alpha\mathbf{y}_k)\Big).$$

The minimization over α'_d yields $\alpha'_d = (\mathbf{x}'_d{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$ for all $d = 1, \ldots, D$. This gives $\alpha' = (\mathbf{x}'{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$ hence $\alpha = (\mathbf{x}{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$, which remarkably does *not* depend on Σ.
The minimized value of the distinguisher is thus

$$\begin{aligned}\min_{\alpha}\ \mathrm{tr}\Big((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\Big) &= \mathrm{tr}\Big((\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)\Big)\\ &= \mathrm{tr}\Big((\mathrm{Id} - {\mathbf{y}_k}^{\mathrm{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1})^2\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\Big)\\ &= \mathrm{tr}\Big(\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\Big) - \mathrm{tr}\Big({\mathbf{y}_k}^{\mathrm{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}\ \mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\Big)\end{aligned}$$

where Id is the $D \times D$ identity matrix and $\mathrm{tr}\Big(\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\Big)$ is a constant independent of k. □

Proof.

Let $\mathbf{x}' = \Sigma^{-1/2}\,\mathbf{x}$ and $\alpha' = \Sigma^{-1/2}\,\alpha$.
The optimal distinguisher minimizes the following expression over $\alpha \in \mathbb{R}^{D\times S}$:

$$\mathrm{tr}\Big((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\Big) = \mathrm{tr}\Big((\mathbf{x}' - \alpha'\mathbf{y}_k)^{\mathsf{T}}(\mathbf{x}' - \alpha'\mathbf{y}_k)\Big).$$

The minimization over α'_d yields $\alpha'_d = (\mathbf{x}'_d{\mathbf{y}_k}^{\mathsf{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathsf{T}})^{-1}$ for all $d = 1, \ldots, D$. This gives $\alpha' = (\mathbf{x}'{\mathbf{y}_k}^{\mathsf{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathsf{T}})^{-1}$ hence $\alpha = (\mathbf{x}{\mathbf{y}_k}^{\mathsf{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathsf{T}})^{-1}$, which remarkably does *not* depend on Σ.
The minimized value of the distinguisher is thus

$$\begin{aligned}\min_{\alpha}\ \mathrm{tr}\Big((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\Big) &= \mathrm{tr}\Big((\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)\Big)\\ &= \mathrm{tr}\Big((\mathrm{Id} - {\mathbf{y}_k}^{\mathsf{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathsf{T}})^{-1})^2\mathbf{x}^{\mathsf{T}}\Sigma^{-1}\mathbf{x}\Big)\\ &= \mathrm{tr}\Big(\mathbf{x}^{\mathsf{T}}\Sigma^{-1}\mathbf{x}\Big) - \mathrm{tr}\Big({\mathbf{y}_k}^{\mathsf{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathsf{T}})^{-1}\ \mathbf{x}^{\mathsf{T}}\Sigma^{-1}\mathbf{x}\Big)\end{aligned}$$

where Id is the $D \times D$ identity matrix and $\mathrm{tr}\Big(\mathbf{x}^{\mathsf{T}}\Sigma^{-1}\mathbf{x}\Big)$ is a constant independent of k. □

Proof.

Let $\mathbf{x}' = \Sigma^{-1/2}\,\mathbf{x}$ and $\alpha' = \Sigma^{-1/2}\,\alpha$.
The optimal distinguisher minimizes the following expression over $\alpha \in \mathbb{R}^{D\times S}$:

$$\mathrm{tr}\left((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\right) = \mathrm{tr}\left((\mathbf{x}' - \alpha'\mathbf{y}_k)^{\mathrm{T}}(\mathbf{x}' - \alpha'\mathbf{y}_k)\right) = \sum_{d=1}^{D} \|\mathbf{x}' - \alpha'_d\mathbf{y}_k\|^2.$$

The minimization over α'_d yields $\alpha'_d = (\mathbf{x}'_d{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$ for all $d = 1, \ldots, D$. This gives $\alpha' = (\mathbf{x}'{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$ hence $\alpha = (\mathbf{x}{\mathbf{y}_k}^{\mathrm{T}})(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}$, which remarkably does *not* depend on Σ.
The minimized value of the distinguisher is thus

$$\begin{aligned}\min_{\alpha}\ \mathrm{tr}\left((\mathbf{x} - \alpha\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\right) &= \mathrm{tr}\left((\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \alpha^{\mathrm{opt}}\mathbf{y}_k)\right)\\ &= \mathrm{tr}\left((\mathrm{Id} - {\mathbf{y}_k}^{\mathrm{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1})^2\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right)\\ &= \mathrm{tr}\left(\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right) - \mathrm{tr}\left({\mathbf{y}_k}^{\mathrm{T}}(\mathbf{y}_k{\mathbf{y}_k}^{\mathrm{T}})^{-1}\,\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right)\end{aligned}$$

where Id is the $D \times D$ identity matrix and $\mathrm{tr}\left(\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}\right)$ is a constant independent of k. □

Corollary (Alternative Expression of $\mathcal{D}_{\mathrm{ML,sto}}$)

Letting $\mathbf{x}' = \Sigma^{-1/2}\,\mathbf{x}$, *and* $\mathbf{y}'_k = (\mathbf{y}_k \mathbf{y}_k{}^{\mathsf{T}})^{-1/2}\,\mathbf{y}_k$ *as in the proof of Theorem 2, we have*

$$\mathcal{D}_{\mathrm{ML,sto}}(\mathbf{x}, \mathbf{t}) = \underset{k \in \mathbb{F}_2^n}{\operatorname{argmax}} \; \|\mathbf{x}' \mathbf{y}'_k{}^{\mathsf{T}}\|_F. \tag{8}$$

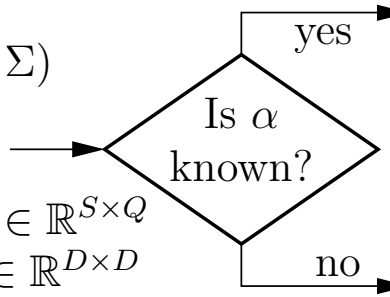
Here the Frobenius norm is of a $D \times S$ *matrix.*

Summary for $S > 2$ Models

Mathematical expression for multivariate ($D \geq 1$) optimal attacks with a linear combination of models ($S \geq 1$):

Leakage model:

$\mathbf{x} = \alpha \mathbf{y}^\star + \mathbf{n}$
$\forall q, n_q \sim \mathcal{N}(0, \Sigma)$
$\mathbf{y}^\star = \phi(\mathbf{t}, k^\star)$
$\mathbf{y}_k = \phi(\mathbf{t}, k)$
$\mathbf{x} \in \mathbb{R}^{D \times Q}, \mathbf{y}_k \in \mathbb{R}^{S \times Q}$
$\alpha \in \mathbb{R}^{D \times S}, \Sigma \in \mathbb{R}^{D \times D}$



Optimal distinguisher:

$\mathcal{D}_{ML}(\mathbf{x}, \mathbf{t}) = \text{argmin}_k \text{ tr}\left((\mathbf{x} - \alpha\mathbf{y}_k)^\mathsf{T}\Sigma^{-1}(\mathbf{x} - \alpha\mathbf{y}_k)\right)$

$\mathcal{D}_{ML,sto}(\mathbf{x}, \mathbf{t}) = \text{argmax}_k \text{ tr}\left(\mathbf{y}_k^\mathsf{T}(\mathbf{y}_k\mathbf{y}_k^\mathsf{T})^{-1}\mathbf{y}_k \ \mathbf{x}^\mathsf{T}\Sigma^{-1}\mathbf{x}\right)$

input : $\mathbf{x}$, $\mathbf{t}$
output : $\mathcal{D}_{\mathrm{ML}}(\mathbf{x}, \mathbf{t})$

// Initialize to zero a matrix $x'_{d,t}$ of size $D \times 2^n$
// Initialize to zero a vector n_t of length 2^n
for $q \in \{1, \ldots, Q\}$ **do**
    $x'_{t_q} \leftarrow x'_{t_q} + \Sigma^{-1/2} x_q$
    $n_{t_q} \leftarrow n_{t_q} + 1$

return $\mathrm{argmin}_{k \in \mathcal{K}} \sum_{d=1}^{D} \sum_t -2 x'_t \alpha'_d y(t,k) + n_t (\alpha'_d y(t,k))^2$

Algorithm 1: Fast computation algorithm for $\mathcal{D}_{\mathrm{ML}}$

input : $\mathbf{x}$, $\mathbf{t}$
output : $\mathcal{D}_{\mathrm{ML,sto}}(\mathbf{x}, \mathbf{t})$

// Precompute the $\#\mathcal{K} = 2^n$ matrices $y'(k)$ of size $S \times 2^n$, such that $y'(k) = (\frac{1}{2^n} \sum_t y(t,k) y(t,k)^{\mathsf{T}})^{-1/2} y(k)$. Note that there is only one matrix if the EIS holds [SLP05, Def. 2]

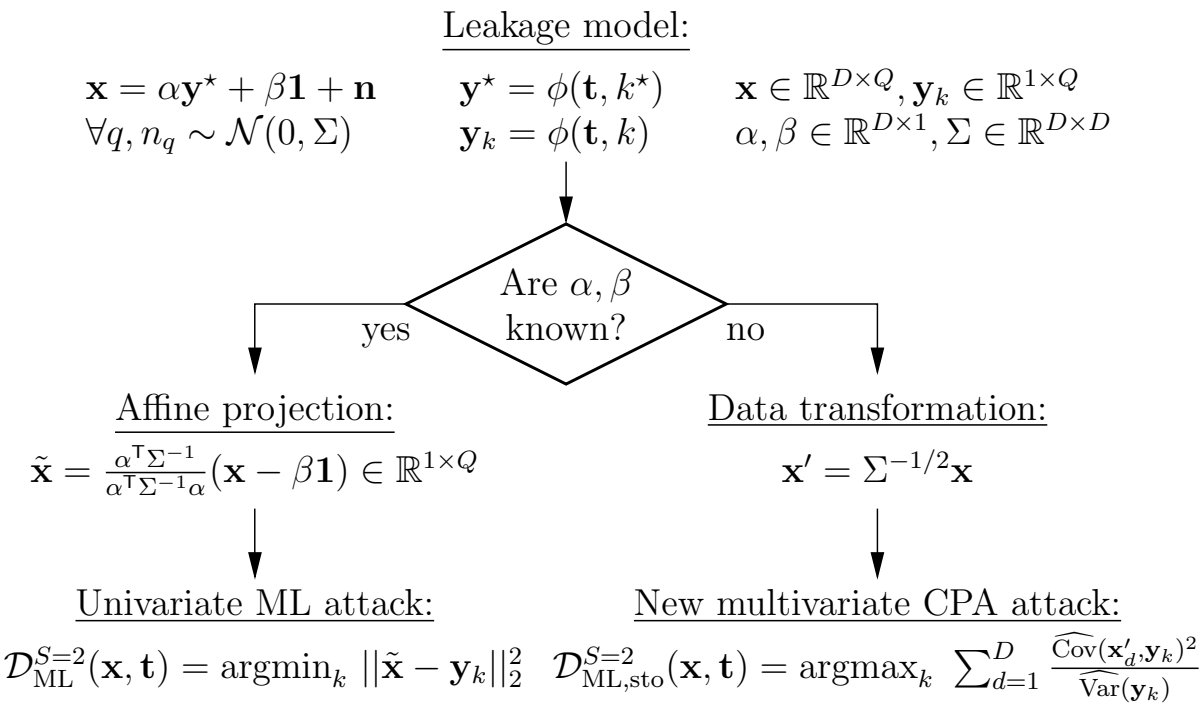
// Initialize to zero a matrix $x'_{d,t}$ of size $D \times 2^n$
for $q \in \{1, \ldots, Q\}$ **do**
  $x'_{t_q} \leftarrow x'_{t_q} + \Sigma^{-1/2} x_q$ // In-place accumulation of a row in matrix x'

return $\mathrm{argmax}_{k \in \mathcal{K}} \ \|x' y'(k)^{\mathsf{T}}\|_F$

Algorithm 2: Fast computation algorithm for $\mathcal{D}_{\mathrm{ML,sto}}$ when $\mathbf{t}$ is balanced

Summary for $S = 2$ Models .... (extension of [BGH+15])

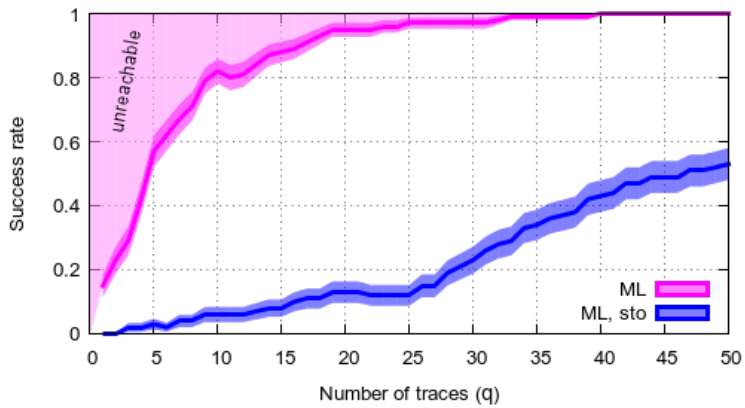
Modus operandi for multivariate ($D \geq 1$) optimal attacks with one model **Y** associated to envelope $\alpha \in \mathbb{R}^{D\times 1}$ and a constant offset $\beta \in \mathbb{R}^{D\times 1}$ ($S = 2$):



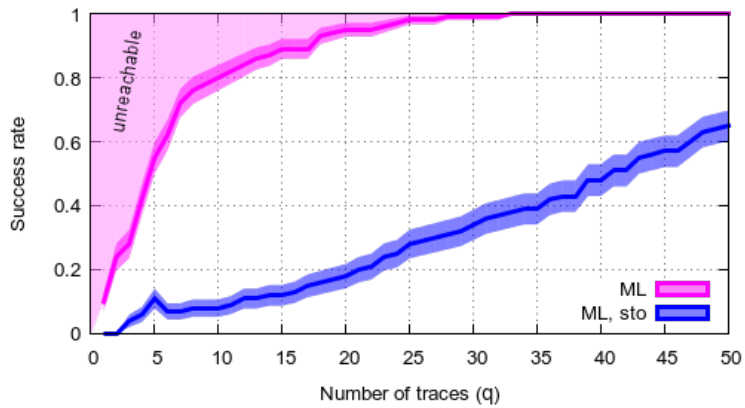
Presentation Outline

Simulations for $D = 3$, $S = 5$, $n = 4$, $\sigma = 1$ (AR noise with $\rho = 0.5$).

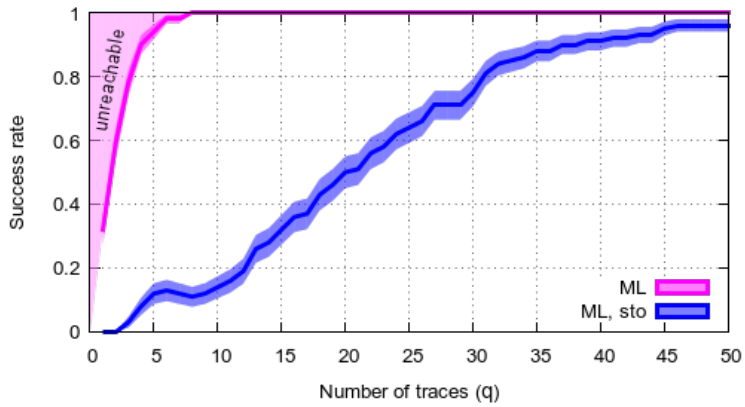
α identical and Σ isotropic



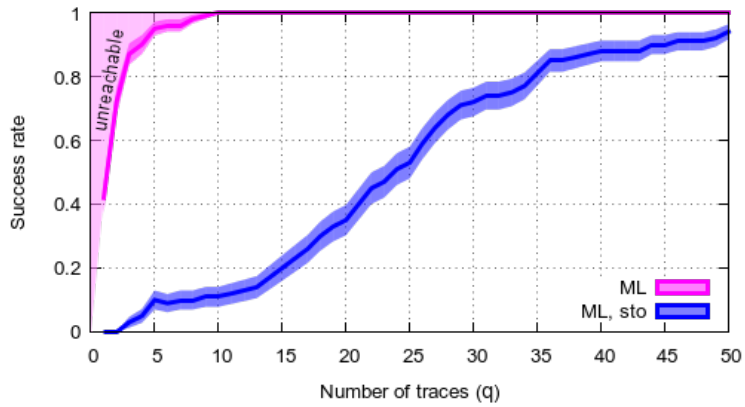
α identical and Σ auto-regressive



α proportional and Σ isotropic

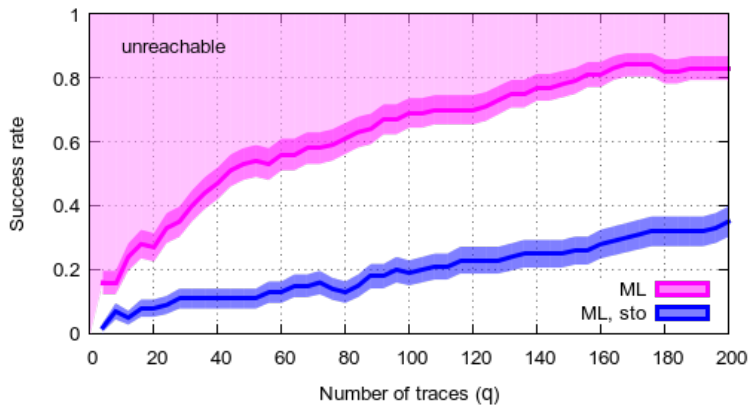


α proportional and Σ auto-regressive

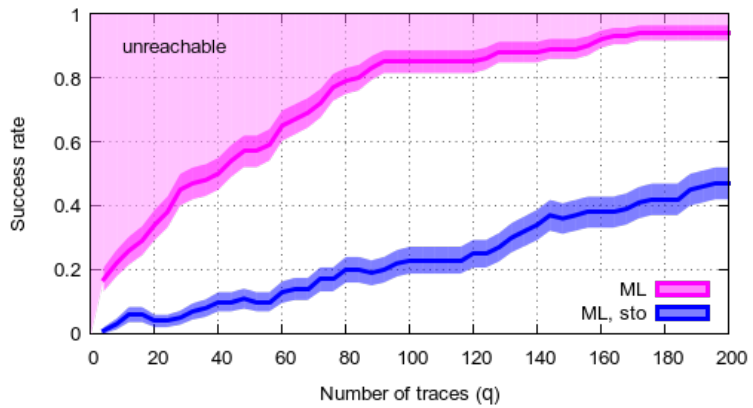


Simulations for $D = 3$, $S = 5$, $n = 4$, $\sigma = 4$ (AR noise with $\rho = 0.5$).

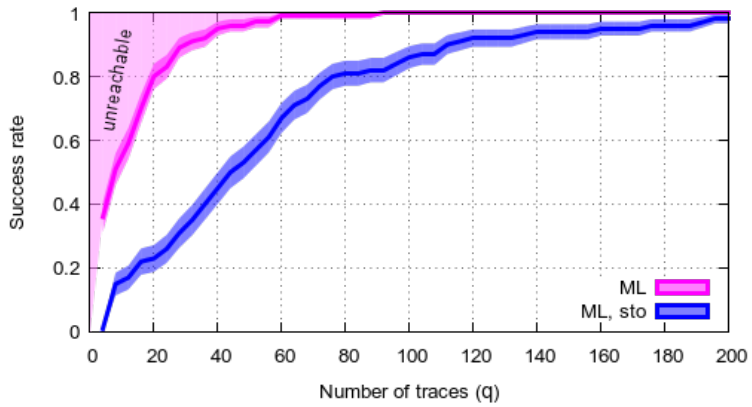
α identical and Σ isotropic



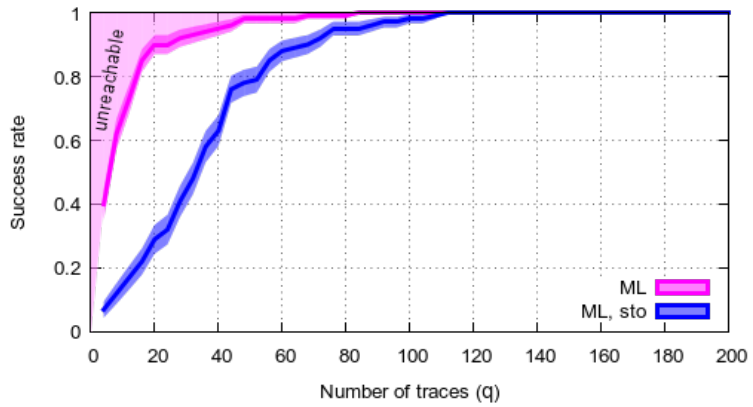
α identical and Σ auto-regressive



α proportional and Σ isotropic



α proportional and Σ auto-regressive



Real-world traces

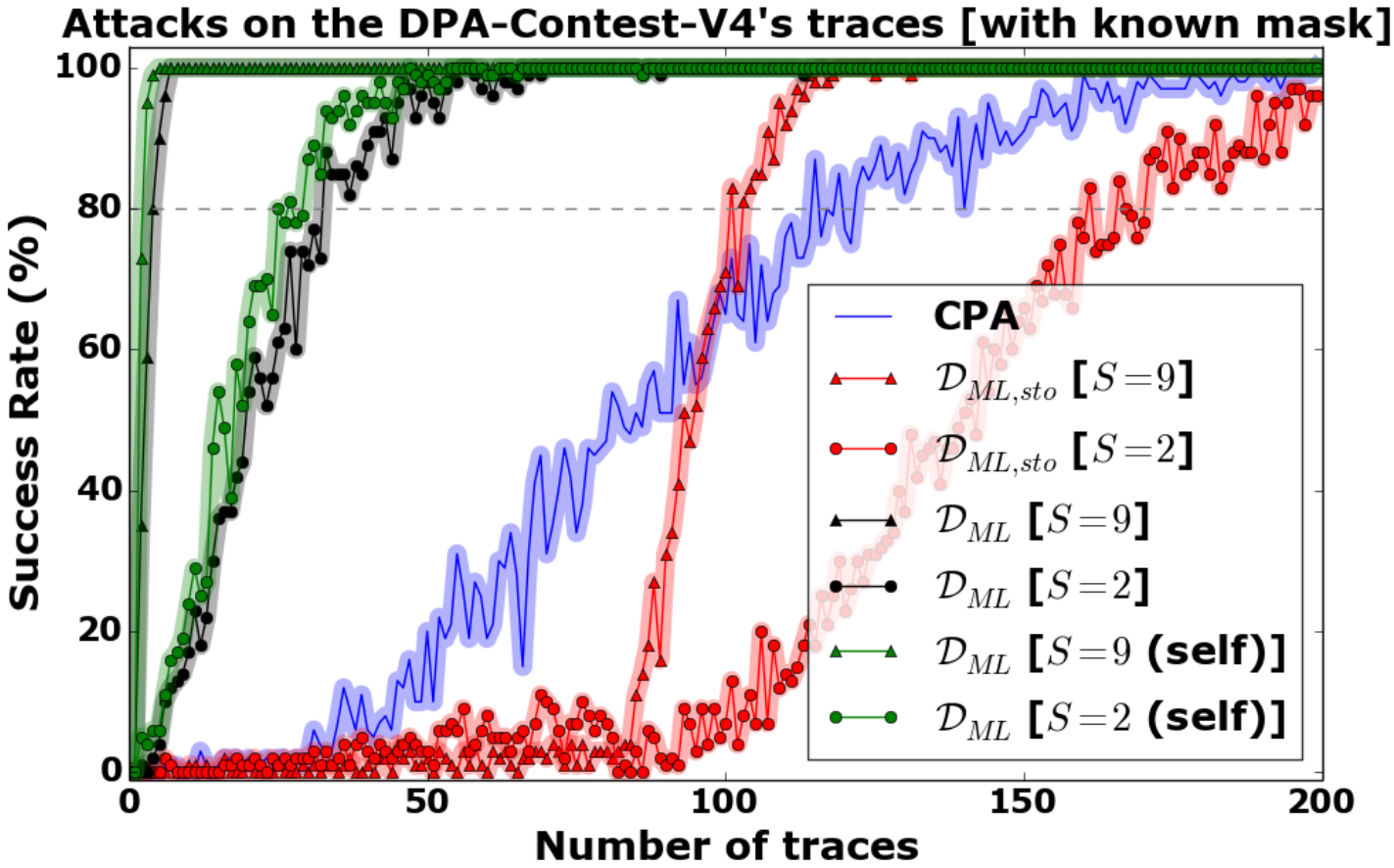


Figure 1 : Comparison of success rate of CPA, $\mathcal{D}_{\mathrm{ML,sto}}$ for $S \in \{9, 2\}$, and $\mathcal{D}_{\mathrm{ML}}$ for $S \in \{9, 2\}$ (with two distinct learning methods)

Presentation Outline

Perspectives?

- First-order success exponent E (recall: $\mathrm{SR} = 1 - \exp -qE$) for:
  - $\mathcal{D}_{\mathrm{ML}}$ (for which $E \approx \frac{1}{2} \cdot \mathrm{SNR} \cdot \min_{k \neq k^\star} \kappa_{k,k^\star}$ — see [GHR15, Proposition 5]) and
  - $\mathcal{D}_{\mathrm{ML,sto}}$ (**TBD**)

  would allow to *quantity* the loss of online profiling (formal analysis of "Templates vs. Stochastic Methods" by Gierlichs, Lemke-Rust and Paar at CHES 2006 [GLRP06]).
- Same research direction to determine the dimensionality S of the basis? $S > n$ includes non-linear leakage (combination of bits).

Optimal Attacks for Multivariate and Multimodel Side-Channel Leakages

Nicolas Bruneau, Sylvain Guilley, Annelie Heuser,
Damien Marion and Olivier Rioul

Saturday August 20, 2016

PROOFS, UCSB, Santa Barbara

[BGH+15] Nicolas Bruneau, Sylvain Guilley, Annelie Heuser, Damien Marion, and Olivier Rioul.
Less is More - Dimensionality Reduction from a Theoretical Perspective.
In Tim Güneysu and Helena Handschuh, editors, *Cryptographic Hardware and Embedded Systems - CHES 2015 - 17th International Workshop, Saint-Malo, France, September 13-16, 2015, Proceedings*, volume 9293 of *Lecture Notes in Computer Science*, pages 22–41. Springer, 2015.

[GHR15] Sylvain Guilley, Annelie Heuser, and Olivier Rioul.
A Key to Success - Success Exponents for Side-Channel Distinguishers.
In Alex Biryukov and Vipul Goyal, editors, *Progress in Cryptology - INDOCRYPT 2015 - 16th International Conference on Cryptology in India, Bangalore, India, December 6-9, 2015, Proceedings*, volume 9462 of *Lecture Notes in Computer Science*, pages 270–290. Springer, 2015.

[GLRP06] Benedikt Gierlichs, Kerstin Lemke-Rust, and Christof Paar.
Templates vs. Stochastic Methods.
In *CHES*, volume 4249 of *LNCS*, pages 15–29. Springer, October 10-13 2006. Yokohama, Japan.

[HRG14] Annelie Heuser, Olivier Rioul, and Sylvain Guilley.
Good Is Not Good Enough - Deriving Optimal Distinguishers from Communication Theory.

In Lejla Batina and Matthew Robshaw, editors, *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings*, volume 8731 of *Lecture Notes in Computer Science*, pages 55–74. Springer, 2014.

[SLP05] Werner Schindler, Kerstin Lemke, and Christof Paar.
A Stochastic Model for Differential Side Channel Cryptanalysis.
In LNCS, editor, *CHES*, volume 3659 of *LNCS*, pages 30–46. Springer, Sept 2005.
Edinburgh, Scotland, UK.

[TEL14] TELECOM ParisTech SEN research group.
DPA Contest (4^{th} edition), 2013–2014.
http://www.DPAcontest.org/v4/.