

Efficient SPA Countermeasures using Redundant Number Representation with Application to ML-KEM

Rishub Nagpal¹ Vedad Hadžić² Robert Primas² Stefan Mangard¹

¹Graz University of Technology | ²Intel Labs

SAC 2025

> isec.tugraz.at

- Implementation security of PQC against worst-case side-channel attacks such as SPA and SASCA
- Analyze Redundant Number Representation (RNR) as a countermeasure against SPA
 - Mutual Information Analysis of RNR for arbitrary integer ring sizes.
 - Application of RNR to ML-KEM resulting in 62.8% overhead for the NTTT

 - W.J. Demonstrate countermeasure effectiveness in both against the strongeside known SPA affact on MI-KEM

- Implementation security of PQC against worst-case side-channel attacks such as SPA and SASCA
- Analyze Redundant Number Representation (RNR) as a countermeasure against SPA
 - Mutual Information Analysis of RNR for arbitrary integer ring sizes.
 - Application of RNR to ML-KEM resulting in **62.8**% **overhead for the NTT** and **0**% **overhead for the INTT**.
 - **Demonstrate countermeasure effectiveness** in both against the strongest known SPA attack on ML-KEM.

- Implementation security of PQC against worst-case side-channel attacks such as SPA and SASCA
- Analyze Redundant Number Representation (RNR) as a countermeasure against SPA
 - Mutual Information Analysis of RNR for arbitrary integer ring sizes.
 - Application of RNR to ML-KEM resulting in **62.8**% **overhead for the NTT** and **0**% **overhead for the INTT**.
 - **Demonstrate countermeasure effectiveness** in both against the strongest known SPA attack on ML-KEM.

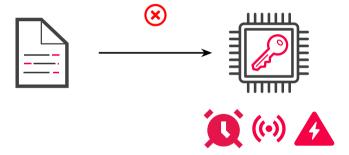
- Implementation security of PQC against worst-case side-channel attacks such as SPA and SASCA
- Analyze Redundant Number Representation (RNR) as a countermeasure against SPA
 - Mutual Information Analysis of RNR for arbitrary integer ring sizes.
 - Application of RNR to ML-KEM resulting in **62.8% overhead for the NTT** and **0% overhead for the INTT**.
 - **Demonstrate countermeasure effectiveness** in both against the strongest known SPA attack on ML-KEM.

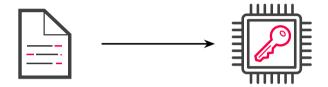
- Implementation security of PQC against worst-case side-channel attacks such as SPA and SASCA
- Analyze Redundant Number Representation (RNR) as a countermeasure against SPA
 - Mutual Information Analysis of RNR for arbitrary integer ring sizes.
 - Application of RNR to ML-KEM resulting in **62.8% overhead for the NTT** and **0% overhead for the INTT**.
 - **Demonstrate countermeasure effectiveness** in both against the strongest known SPA attack on ML-KEM.

Motivation isec.tugraz.at









The Side-channel Problem

Cryptographic algorithms can be secure from a "black box" view, but insecure when implemented in the real-world due to physical effects.

 Kyber and Dilithium are standardized by NIST as ML-KEM and ML-DSA.



 Kyber and Dilithium are standardized by NIST as ML-KEM and ML-DSA.



Side-channel attacks are still a problem despite quantum resistance...

Chosen Ciphertext *k*-trace attack of Hamburg et al. [Ham+21]

- Chosen Ciphertext k-trace attack of Hamburg et al. [Ham+21]
 - One of the strongest known attacks on ML-KEM.

- Chosen Ciphertext k-trace attack of Hamburg et al. [Ham+21]
 - One of the strongest known attacks on ML-KEM.
 - Possible with only a few measurements via Soft-analytical Side-channel Analysis (SASCA).

- Chosen Ciphertext k-trace attack of Hamburg et al. [Ham+21]
 - One of the strongest known attacks on ML-KEM.
 - Possible with only a few measurements via Soft-analytical Side-channel Analysis (SASCA).
 - Even against CCA2-secure masked implementations...

k-trace attack of Hamburg et al. [Ham+21]

ML-KEM.PKE Decryption

```
Input: ciphertext c = (c_1, c_2), sk = \hat{s}

Output: message m \in \mathcal{R}_q

1: (u, v) = (\text{Decompress}(c_1), \text{Decompress}(c_2))

2: return m = v - \text{NTT}^{-1}(\hat{s}^T \circ \text{NTT}(u))
```

k-trace attack of Hamburg et al. [Ham+21]

ML-KEM.PKE Decryption

```
Input: ciphertext c = (c_1, c_2), sk = \hat{s}
Output: message m \in \mathcal{R}_q
1: (u, v) = (\text{Decompress}(c_1), \text{Decompress}(c_2))
2: return m = v - \text{NTT}^{-1}(\hat{s}^T \circ \text{NTT}(u))
sparse product
```

k-trace attack of Hamburg et al. [Ham+21]

ML-KEM.PKE Decryption

```
Input: ciphertext c = (c_1, c_2), sk = \hat{s}

Output: message m \in \mathcal{R}_q

1: (u, v) = (\text{Decompress}(c_1), \text{Decompress}(c_2))

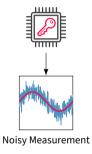
2: return m = v - \text{NTT}^{-1}(\hat{s}^T \circ \text{NTT}(u))

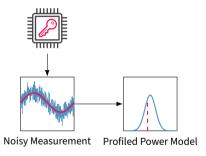
sparse product
```

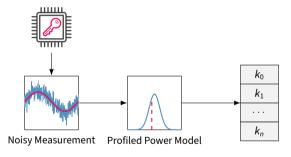
Chosen Ciphertext k-trace attack

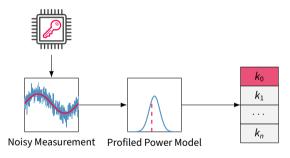
Chosen ciphertexts enable divide-and-conquer recovery of \$\frac{1}{2}\$ from the NTT-1 of the sparse product.

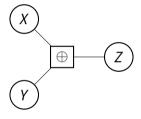


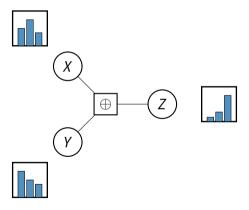


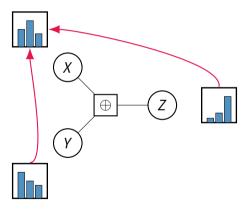


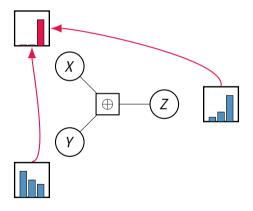












```
Input: \hat{f} \in \mathbb{Z}_a^{256}
Output: f \in \mathbb{Z}_q^{256}
  1. f \leftarrow \hat{f}
  2: k \leftarrow 127; j \leftarrow 0
  3: for len \leftarrow 2: len < 128: len \leftarrow 2 \cdot len do
              for start \leftarrow 0; start < 256; start \leftarrow i + \text{len do}
  4:
                     for i \leftarrow \text{start}: i < \text{start} + \text{len}: i + + \text{do}
  5:
  6:
                         f_i \leftarrow \mathsf{barrett\_reduce}(t + f_{i+\mathsf{len}})
  7:
                          f_{i+\text{len}} \leftarrow f_{i+\text{len}} - t
                    f_{i+\text{len}} \leftarrow \text{montgomery\_reduce}(\zeta^k \cdot f_{i+\text{len}})
                     k \leftarrow k - 1
 10:
```

```
Input: \hat{f} \in \mathbb{Z}_a^{256}
Output: f \in \mathbb{Z}_q^{256}
  1. f \leftarrow \hat{f}
  2: k \leftarrow 127; j \leftarrow 0
  3: for len \leftarrow 2: len < 128: len \leftarrow 2 \cdot len do
              for start \leftarrow 0; start < 256; start \leftarrow i + \text{len do}
  4:
                     for i \leftarrow \text{start}: i < \text{start} + \text{len}: i + + \text{do}
  5:
  6:
                          f_i \leftarrow \mathsf{barrett\_reduce}(t + f_{i+\mathsf{len}})
  7:
                          f_{i+\text{len}} \leftarrow f_{i+\text{len}} - t
                     f_{i+\text{len}} \leftarrow \text{montgomery\_reduce}(\zeta^k \cdot f_{i+\text{len}})
                     k \leftarrow k - 1
 10:
```

⊳ 7 layers

```
Input: \hat{f} \in \mathbb{Z}_a^{256}
Output: f \in \mathbb{Z}_q^{256}
  1. f \leftarrow \hat{f}
  2: k \leftarrow 127; j \leftarrow 0
  3: for len \leftarrow 2: len < 128: len \leftarrow 2 \cdot len do
              for start \leftarrow 0; start < 256; start \leftarrow i + \text{len do}
  4:
                     for i \leftarrow \text{start}: i < \text{start} + \text{len}: i + + \text{do}
  5:
  6:
                          f_i \leftarrow \mathsf{barrett\_reduce}(t + f_{i+\mathsf{len}})
  7:
                          f_{i+\text{len}} \leftarrow f_{i+\text{len}} - t
                     f_{i+\text{len}} \leftarrow \text{montgomery\_reduce}(\zeta^k \cdot f_{i+\text{len}})
                     k \leftarrow k - 1
 10:
```

⊳ 7 layers ⊳ 256 coeffs

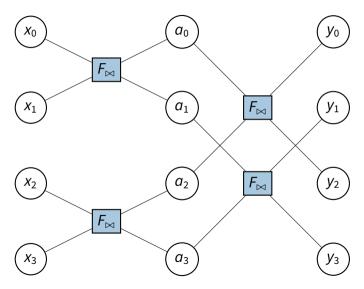
```
Input: \hat{f} \in \mathbb{Z}_a^{256}
Output: f \in \mathbb{Z}_q^{256}
  1. f \leftarrow \hat{f}
  2: k \leftarrow 127; j \leftarrow 0
  3: for len \leftarrow 2: len < 128: len \leftarrow 2 \cdot len do

⊳ 7 lavers

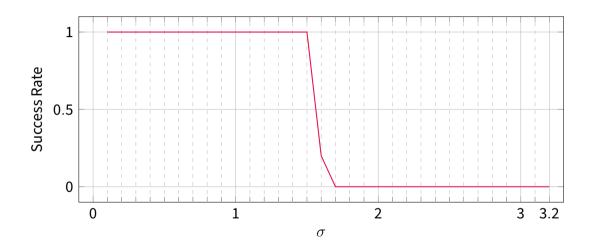
              for start \leftarrow 0; start < 256; start \leftarrow i + \text{len do}
                                                                                                                                > 256 coeffs.
  4:
                    for i \leftarrow \text{start}: i < \text{start} + \text{len}: i + + \text{do}
                                                                                                                   ⊳ Select coeff. pairs
  5:
  6:
                         f_i \leftarrow \mathsf{barrett\_reduce}(t + f_{i+\mathsf{len}})
  7:
                         f_{i+\text{len}} \leftarrow f_{i+\text{len}} - t
                    f_{i+\text{len}} \leftarrow \text{montgomery\_reduce}(\zeta^k \cdot f_{i+\text{len}})
                    k \leftarrow k - 1
 10:
```

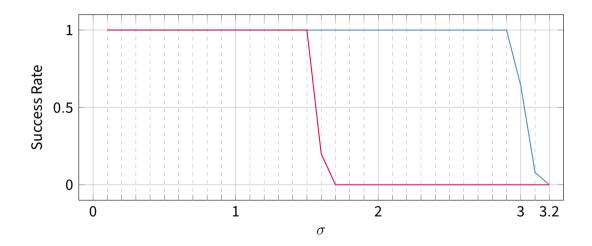
```
Input: \hat{f} \in \mathbb{Z}_a^{256}
Output: f \in \mathbb{Z}_q^{256}
  1. f \leftarrow \hat{f}
  2: k \leftarrow 127; j \leftarrow 0
  3: for len \leftarrow 2: len < 128: len \leftarrow 2 \cdot len do
              for start \leftarrow 0; start < 256; start \leftarrow i + \text{len do}
  4:
                     for i \leftarrow \text{start}: i < \text{start} + \text{len}: i + + \text{do}
  5:
                            t \leftarrow f_i
  6:
                           f_i \leftarrow \mathsf{barrett\_reduce}(t + f_{i+\mathsf{len}})
  7:
                           f_{i+\text{len}} \leftarrow f_{i+\text{len}} - t
                      f_{i+\text{len}} \leftarrow \text{montgomery\_reduce}(\zeta^k \cdot f_{i+\text{len}})
                     k \leftarrow k - 1
 10:
```

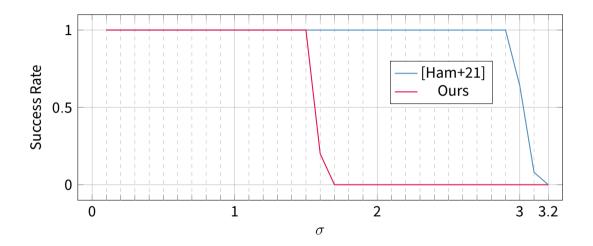
```
⊳ 7 layers
⊳ 256 coeffs.
⊳ Select coeff. pairs
⊳ GS-Butterfly
```



$$F_{oxdotsim}(x_0,x_1,y_0,y_1) = egin{cases} 1 & y_0 = x_0 + \zeta x_1 mod q \land \ y_1 = x_0 - \zeta x_1 mod q \ 0 & ext{otherwise} \end{cases}$$







• Operates in a Polynomial ring with coefficients in \mathbb{Z}_q ; q=3329

- Operates in a Polynomial ring with coefficients in \mathbb{Z}_q ; q=3329
- $\log_2 q \approx 11.7$ -bits \odot stored in 16-bit machine representations.

- Operates in a Polynomial ring with coefficients in \mathbb{Z}_q ; q=3329
- $\log_2 q \approx 11.7$ -bits \bigcirc stored in 16-bit machine representations.
- Efficient implementations represent the integers in the signed range $\left[\left|\frac{-q}{2}\right|,\left\lceil\frac{q}{2}\right\rceil\right)$

 The signed representation of the Polynomial ring makes Barrett and Montgomery reductions more efficient.

- The signed representation of the Polynomial ring makes Barrett and Montgomery reductions more efficient.
- Great performance optimization! Less instructions, enables lazy reductions etc.

- The signed representation of the Polynomial ring makes Barrett and Montgomery reductions more efficient.
- Great performance optimization! Less instructions, enables lazy reductions etc.
- Sounds great, but...

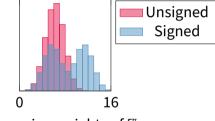
- The signed representation of the Polynomial ring makes Barrett and Montgomery reductions more efficient.
- Great performance optimization! Less instructions, enables lazy reductions etc.
- Sounds great, but...

- The signed representation of the Polynomial ring makes Barrett and Montgomery reductions more efficient.
- Great performance optimization! Less instructions, enables lazy reductions etc.
- Sounds great, but...

Side-channel Distinguisher [TMS24]

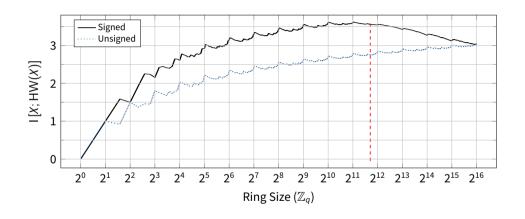
Small integer ranges (relative to the machine-word size) will have a large Hamming weight disparity between positive and negative numbers.

Hamming Weight Distributions of \mathbb{Z}_a



Hamming weights of \mathbb{Z}_a

Mutual Information Analysis of Machine Representations



Input polynomial: $\mathbf{X} = (X_i)_{i=0}^{255}$ where $X_i \in \mathbb{Z}_q$.

- Input polynomial: $\mathbf{X} = (X_i)_{i=0}^{255}$ where $X_i \in \mathbb{Z}_q$.
- **2** Total Entropy: $H[X] = \sum_{X} H[X_i] \approx 2995.4 \, \text{bits.}$

- Input polynomial: $\mathbf{X} = (X_i)_{i=0}^{255}$ where $X_i \in \mathbb{Z}_q$.
- Total Entropy: $H[X] = \sum_{X} H[X_i] \approx 2995.4 \text{ bits.}$
- Adversary must learn: $H[X] \sum_{X} I[X_i; W(X_i)]$

- Input polynomial: $\mathbf{X} = (X_i)_{i=0}^{255}$ where $X_i \in \mathbb{Z}_q$.
- Total Entropy: $H[X] = \sum_{X} H[X_i] \approx 2995.4 \text{ bits.}$
- Adversary must learn: $H[X] \sum_{X} I[X_i; W(X_i)]$

± Signed

 \sim 2083.8 bits.

- Input polynomial: $\mathbf{X} = (X_i)_{i=0}^{255}$ where $X_i \in \mathbb{Z}_q$.
- Total Entropy: $H[X] = \sum_{\mathbf{x}} H[X_i] \approx 2995.4 \, \text{bits.}$
- Adversary must learn: $H[X] \sum_{\mathbf{x}} I[X_i; W(X_i)]$

± Signed

 $\approx 2083.8 \, \mathrm{bits}$

Unsigned

 \approx 2289 9 bits.

- Input polynomial: $\mathbf{X} = (X_i)_{i=0}^{255}$ where $X_i \in \mathbb{Z}_q$.
- **2** Total Entropy: $H[X] = \sum_{X} H[X_i] \approx 2995.4 \text{ bits.}$
- Adversary must learn: $H[X] \sum_{X} I[X_i; W(X_i)]$

± Signed

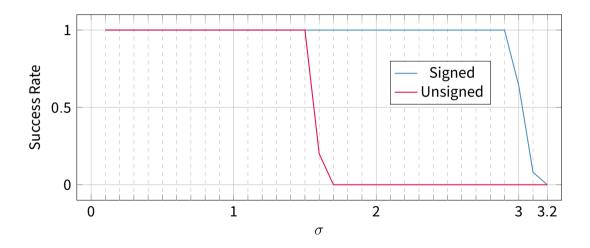
 \sim 2083.8 bits.

+ Unsigned

 \sim 2289.9 bits.

Adversary learns $\approx 206.092 \, \mathrm{bits}$ just from signed representation!

k-trace Attack on ML-KEM



Redundant Number Representation

Application to ML-KEM





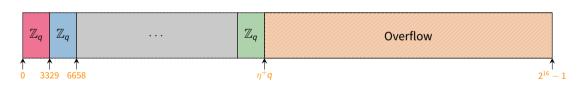
16-bit word



16-bit word



16-bit word



■ Encode $x \in \mathbb{Z}_q$ to $x' \in \mathbb{Z}_{\eta q}$ where x' = x + Kq where K is sampled uniformly from $[0, \eta)$.

- Encode $x \in \mathbb{Z}_q$ to $x' \in \mathbb{Z}_{\eta q}$ where x' = x + Kq where K is sampled uniformly from $[0, \eta)$.
- lacksquare The algorithm operates on η redundant encodings of \mathbb{Z}_q

- Encode $x \in \mathbb{Z}_q$ to $x' \in \mathbb{Z}_{\eta q}$ where x' = x + Kq where K is sampled uniformly from $[0, \eta)$.
- lacksquare The algorithm operates on η redundant encodings of \mathbb{Z}_q
- Ex: $0 \in \mathbb{Z}_q \equiv \{0, 3329, 6658, \dots, (\eta 1)q\}$

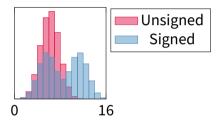
- Encode $x \in \mathbb{Z}_q$ to $x' \in \mathbb{Z}_{\eta q}$ where x' = x + Kq where K is sampled uniformly from $[0, \eta)$.
- lacksquare The algorithm operates on η redundant encodings of \mathbb{Z}_q
- Ex: $0 \in \mathbb{Z}_q \equiv \{0, 3329, 6658, \dots, (\eta 1)q\}$
- Works for signed representations too!

- Encode $x \in \mathbb{Z}_q$ to $x' \in \mathbb{Z}_{\eta q}$ where x' = x + Kq where K is sampled uniformly from $[0, \eta)$.
- The algorithm operates on η redundant encodings of \mathbb{Z}_q
- Ex: $0 \in \mathbb{Z}_q \equiv \{0, 3329, 6658, \dots, (\eta 1)q\}$
- Works for signed representations too!

Outcome

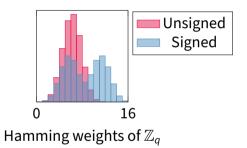
 η encodings means upto η unique Hamming weights for a given x - Makes SPA harder!

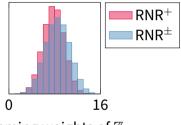
Hamming weight distriubtions of RNR



Hamming weights of \mathbb{Z}_a

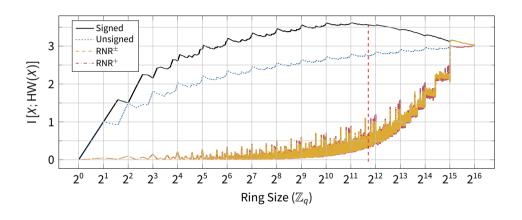
Hamming weight distriubtions of RNR

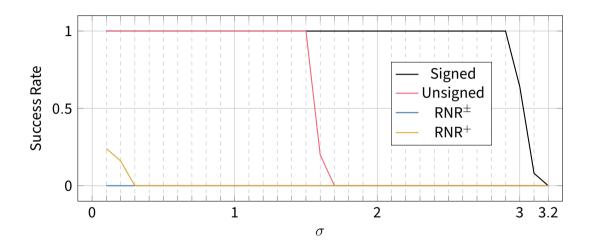




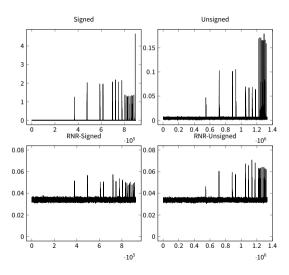
Hamming weights of $\mathbb{Z}_{\eta q}$

Redundant Number Representation

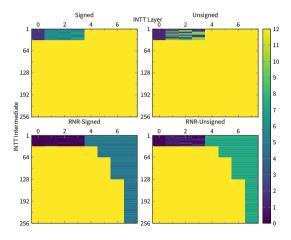




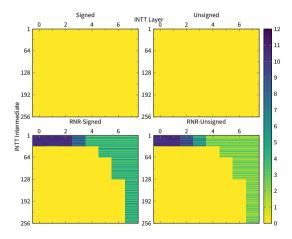
k-trace Attack on a ARM Cortex-M4 - SNR



k-trace Attack on a ARM Cortex-M4 - PI Estimate



k-trace Attack on a ARM Cortex-M4 - SASCA Result



Implementation and Performance Results

Implementation	KCycles (⋅10³)			
	-00	Overhead	-03	Overhead
Signed-NTT	127.02		26.48	
Unsigned-NTT	158.00		36.75	
$RNR^\pm ext{-}NTT$	196.01	42.7%	50.70	62.8%
RNR ⁺ -NTT	260.52	49.0%	84.74	79.0%
Signed-NTT ⁻¹	202.04		42.61	
Unsigned-NTT ⁻¹	270.39		64.91	
$RNR^{\pm} ext{-}NTT^{-1}$	203.19	0.6%	42.61	0%
RNR ⁺ -NTT ⁻¹	305.59	12.2%	91.15	27.7%

Comparison to Shuffling [Rav+20]

			\ /	v				
		KCycles ($\times 10^3$)						
Countermeasures	Shuffle Algo.	Count	Overhead	Shuffle	Rand.			
			(%)					
Kyber NTT								
Unprotected	NA	31.0	-	-	-			
Coarse-Full-Shuffled	Knuth-Yates	87.2	181.1	16.6 (19 %)	38.4 (44.1 %)			
Coarse-In-Group-Shuffle		84.4	$\boldsymbol{172.2}$	17.1~(20.3 %)	32.4~(38.4 %)			
Basic-Fine-Shuffled	Arith. cswap	76.7	147.4	35.1 (45.7 %)	9.5~(12.4 %)			
Bitwise-Fine-Shuffle		142.6	356	100.1~(70.2%)	9.5~(6.7 %)			
Kyber INTT								
Unprotected	NA	50.6	-	-	-			
Coarse-Full-Shuffled	Knuth-Yates	113.3	123.8	16.6 (14.6 %)	38.4 (33.9 %)			
Coarse-In-Group-Shuffled		101.2	99.9	$16 \ (\mathbf{15.8\%})$	33~(32.6 %)			
Basic-Fine-Shuffled	Arith. cswap	101.8	101.1	40.9 (40.1 %)	9.5 (9.4 %)			
Bitwise-Fine-Shuffled		172.4	240.8	$102.2~(\mathbf{59.3\%})$	9.6~(5.5 %)			

Conclusion

- Even small performance optimizations can have unforseen and impactful consequences.
- NR is sufficient at preventing the strongest known SPA attack against ML-KEM.
- Can be achieved with a low performance impact and simple to implement!

- rishub.nagpal@tugraz.at
- https://github.com/rishubn/rnr-kyber-spa



Acknowledgments

This research was funded in whole or in part by the Austrian Science Fund (FWF) (FWF SFB project SPyCoDe 10.55776/F85) and the Austrian Research Promotion Agency (FFG) via the AWARE project (FFG grant number 891092).







Bibliography

- [Ham+21] Mike Hamburg et al. Chosen Ciphertext k-Trace Attacks on Masked CCA2 Secure Kyber. IACR Trans. Cryptogr. Hardw. Embed. Syst. 2021.4 (2021), pp. 88–113. DOI: 10.46586/TCHES.V2021.14.88–113. URL: https://doi.org/10.46586/tches.v2021.i4.88–113.
- Prasanna Ravi et al. On Configurable SCA Countermeasures Against Single Trace Attacks for the NTT A Performance Evaluation Study over Kyber and Dilithium on the ARM Cortex-M4. Security, Privacy, and Applied Cryptography Engineering 10th International Conference, SPACE 2020, Kolkata, India, December 17-21, 2020, Proceedings. Ed. by Lejla Batina, Stjepan Picek, and Mainack Mondal. Vol. 12586. Lecture Notes in Computer Science. Springer, 2020, pp. 123–146. DOI: 10.1007/978-3-030-66626-2_7. URL: https://doi.org/10.1007/978-3-030-66626-2%5C_7.
- [TMS24] Tolun Tosun, Amir Moradi, and Erkay Savas. Exploiting the Central Reduction in Lattice-Based Cryptography. IEEE Access 12 (2024), pp. 166814–166833. DOI: 10.1109/ACCESS.2024.3494593. URL: https://doi.org/10.1109/ACCESS.2024.3494593.

Backup Slides

Derivation of η

$$\eta^+ q + \left(rac{\eta^+ q^2}{2^{16}} + \eta^+ q
ight) + q < 2^{16}$$

$$\eta^+ \cdot \left(2q + rac{q^2}{2^{16}}
ight) < 2^{16} - q$$

$$\eta^+ < rac{2^{32} - 2^{16}q}{2^{17}q + q^2} < 10$$
 (1)

Modeling the ML-KEM NTT

Number Theoretic Transform

An algorithm analogous to the Discrete Fourier Transform (DFT) which allows one to compute the product of two polynomials efficiently.

In ML-KEM:

Modeling the ML-KEM NTT

Number Theoretic Transform

An algorithm analogous to the Discrete Fourier Transform (DFT) which allows one to compute the product of two polynomials efficiently.

- In ML-KEM:
 - Factors degree-256 polynomials with small 128 degree-2 polynomials

$$(x^{256}+1)=\prod_{i=0}^{127}(x^2-\zeta^{2i+1}),$$

where ζ^n is the *n*-th root-of-unity.

Modeling the ML-KEM NTT

$$\mathsf{NTT}(a) = \hat{a} = \hat{a}_0 + \hat{a}_1 x + \dots \hat{a}_{255} x^{255},$$

$$\hat{a}_i = \sum_{j=0}^{127} a_{2j} \zeta^{(2i+1)j} \qquad \text{and} \qquad \hat{a}_{2i+1} = \sum_{j=0}^{127} a_{2j+1} \zeta^{(2i+1)j}.$$

Multiplication of polynomials: $NTT^{-1}(NTT(f) \circ NTT(g))$.