The Design of Falcon

Thomas Prest (joint work with Pierre-Alain Fouque, Jeffrey Hoffstein, Paul Kirchner, Vadim Lyubashevsky, Thomas Pornin, Thomas Ricosset, Gregor Seiler, William Whyte, Zhenfei Zhang)

PQShield

Cryptography Standard Design Seminar (02/07/2023)

Roadmap

PQ SHIELD

This talk:

- ✓ High-level description
- ✓ Side-channel security
- ✓ Deployment
- ✓ Algorithmics (a bit)
- ✗ ~~Implementation~~
- ✗ ~~Cryptanalysis~~

Hash-then-Sign

Initial attempts: NTRUSign (1997), GGHSign (2003)

PQ SHIELD

Keygen(1^λ)

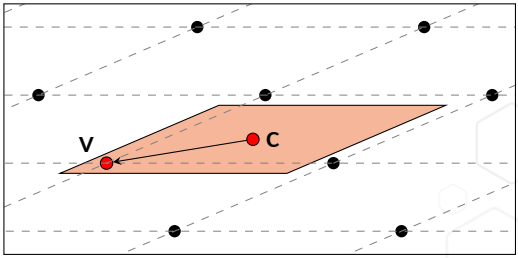
1. Gen. matrices $\mathbf{A}, \mathbf{B}$ such that:
   - $\mathbf{A}$ is pseudorandom
   - $\mathbf{A} \cdot \mathbf{B} = \mathbf{0}$
   - $\mathbf{B}$ has small coefficients
2. $\mathsf{pk} := \mathbf{A}, \mathsf{sk} := \mathbf{B}$

Sign(msg, sk = $\mathbf{B}$)

1. Compute $\mathbf{c}$ such that $\mathbf{A} \cdot \mathbf{c} = H(\mathsf{msg})$
2. $\mathbf{v} := \mathbf{B} \lfloor \mathbf{B}^{-1}\mathbf{c} \rceil$
3. $\mathsf{sig} := \mathbf{s} = (\mathbf{c} - \mathbf{v})$

Verify(msg, pk = $\mathbf{A}$, sig = $\mathbf{s}$)

Check ($\mathbf{s}$ short) & ($\mathbf{A} \cdot \mathbf{s} = H(\mathsf{msg})$)



Initial attempts: NTRUSign (1997), GGHSign (2003)

PQ SHIELD

Keygen(1^λ)

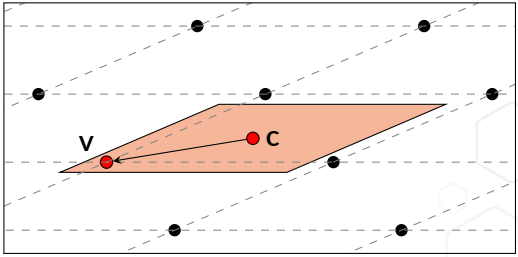
1. Gen. matrices $\mathbf{A}, \mathbf{B}$ such that:
   - $\mathbf{A}$ is pseudorandom
   - $\mathbf{A} \cdot \mathbf{B} = \mathbf{0}$
   - $\mathbf{B}$ has small coefficients
2. $\mathsf{pk} := \mathbf{A}, \mathsf{sk} := \mathbf{B}$

Sign$(\mathsf{msg}, \mathsf{sk} = \mathbf{B})$

1. Compute $\mathbf{c}$ such that $\mathbf{A} \cdot \mathbf{c} = H(\mathsf{msg})$
2. $\mathbf{v} := \mathbf{B} \lfloor \mathbf{B}^{-1}\mathbf{c} \rceil$
3. $\mathsf{sig} := \mathbf{s} = (\mathbf{c} - \mathbf{v})$

Verify$(\mathsf{msg}, \mathsf{pk} = \mathbf{A}, \mathsf{sig} = \mathbf{s})$

Check ($\mathbf{s}$ short) & ($\mathbf{A} \cdot \mathbf{s} = H(\mathsf{msg})$)



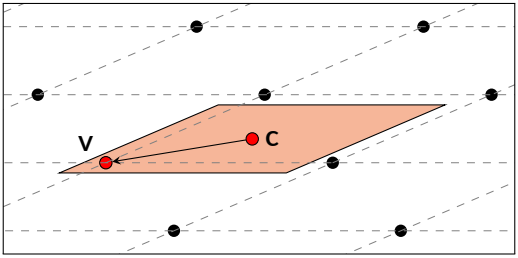
→ **Correctness:** easy

→ **Security:** Finding a short preimage $\mathbf{s}$ of $H(\mathsf{msg})$ should be difficult... or is it?

The parallelepiped attack

Problem: The distribution of the signature **s** is correlated to **B**

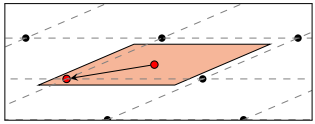
$$\mathbf{s} = \mathbf{c} - \mathbf{B}\left\lfloor \mathbf{B}^{-1}\mathbf{c} \right\rceil \in \left[-\frac{1}{2}, \frac{1}{2}\right) \cdot \mathbf{B} \tag{1}$$



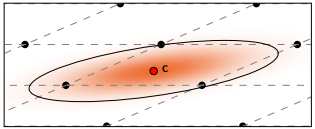
Given many signatures, **B** can be recovered using techniques from Independent Component Analysis (ICA)

- 2006: key-recovery on NTRUSign and GGHSign
- 2012: key-recovery against NTRUSign countermeasures

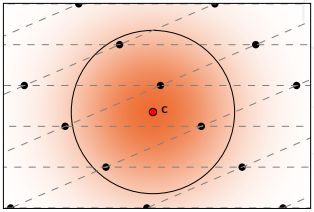
$\mathbf{B}\lfloor \mathbf{B}^{-1}\mathbf{c} \rceil$



$\mathbf{B}\lfloor \mathbf{B}^{-1}\mathbf{c} \rceil_{\sigma_1}$



$\mathbf{B}\left\lfloor \mathbf{B}^{-1}\left(\mathbf{c} + \mathbf{M}\lfloor \mathbf{0} \rceil_{\sigma_2}\right)\right\rceil_{\sigma_1}$



Indistinguishability: For appropriately chosen parameters, the rightmost procedure outputs a distribution close to a perfect Gaussian $D_{\Lambda(\mathbf{B}),\mathbf{c},\sigma}$.

Consequence: these two worlds are indistinguishable (in the ROM)

1. Sample a short vector $\mathbf{s}$, then set $H(\mathbf{msg}) = \mathbf{A} \cdot \mathbf{s}$
2. Compute $H(\mathbf{msg})$, then use $\mathbf{B}$ to sample a short preimage $\mathbf{s}$ of $H(\mathbf{msg})$

What are the next steps?

The GPV framework requires two ingredients:

1. A family of trapdoors $(\mathbf{A}, \mathbf{B})$
2. A trapdoor sampler for computing a short preimage $\mathbf{s}$

Falcon: our goal is to minimize the communication cost

Computing a suitable $(\mathbf{A}, \mathbf{B})$ – NTRU trapdoors

NTRU trapdoors

Let $f, g, F, G \in \mathcal{R}$ such that:

$$fG - gF = q \tag{2}$$

$$h := g/f \bmod q \tag{3}$$

We set $\mathbf{A} = \begin{bmatrix} 1 & h \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} g & G \\ -f & -F \end{bmatrix}$.

Computing a suitable $(\mathbf{A}, \mathbf{B})$ – NTRU trapdoors

NTRU trapdoors

Let $f, g, F, G \in \mathcal{R}$ such that:

$$fG - gF = q \tag{2}$$

$$h := g/f \bmod q \tag{3}$$

We set $\mathbf{A} = \begin{bmatrix} 1 & h \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} g & G \\ -f & -F \end{bmatrix}$.

Pseudorandomness of A: NTRU assumption.

Orthogonality: One can easily show that $\mathbf{A} \cdot \mathbf{B} = \mathbf{0} \bmod q$.

Shortness of B: Given (f, g), one can compute suitable (F, G) such that

$$\left\| \mathrm{Proj}_{\mathrm{Span}(f,g)^{\perp}}(F, G) \right\| \approx \frac{q}{\|(f, g)\|} \tag{4}$$

It is optimal to take $\|(f, g)\| \approx 1.17\sqrt{q}$.

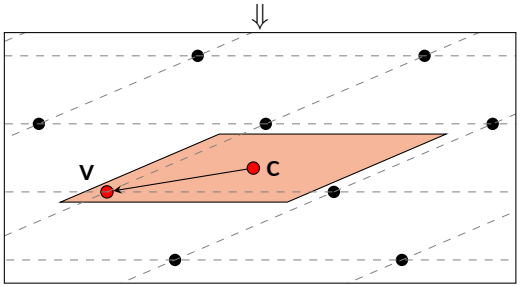
Computing a lattice point v close to the target c

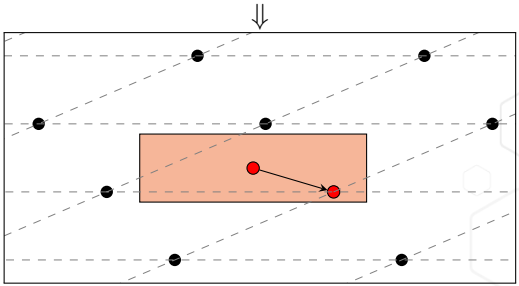
RoundOff$(\mathbf{B}, \mathbf{c})$

1. $\mathbf{t} \leftarrow \mathbf{c} \cdot \mathbf{B}^{-1}$
2. For $j \in \{n, \ldots, 1\}$:
   1. $z_j \leftarrow \lceil t_j \rfloor$
3. Return $\mathbf{v} := \mathbf{z} \cdot \mathbf{B}$

NearestPlane$(\mathbf{B}, \mathbf{L}, \mathbf{c})$

1. $\mathbf{t} \leftarrow \mathbf{c} \cdot \mathbf{B}^{-1}$
2. For $j \in \{n, \ldots, 1\}$:
   1. $z_j \leftarrow \lceil t_j + \sum_{i>j}(t_1 - z_i)L_{i,j} \rfloor$
3. Return $\mathbf{v} := \mathbf{z} \cdot \mathbf{B}$



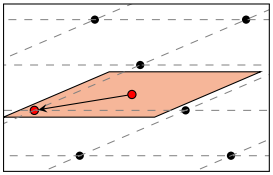


The second algorithm is better for security.

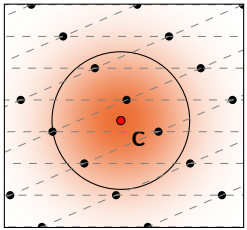
Picking the right σ

We combine NearestPlane with Gaussian rounding to obtain a discretized Gaussian.

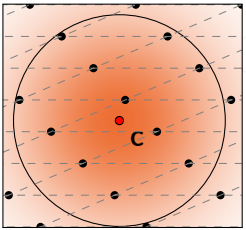
σ too small



The "right" σ

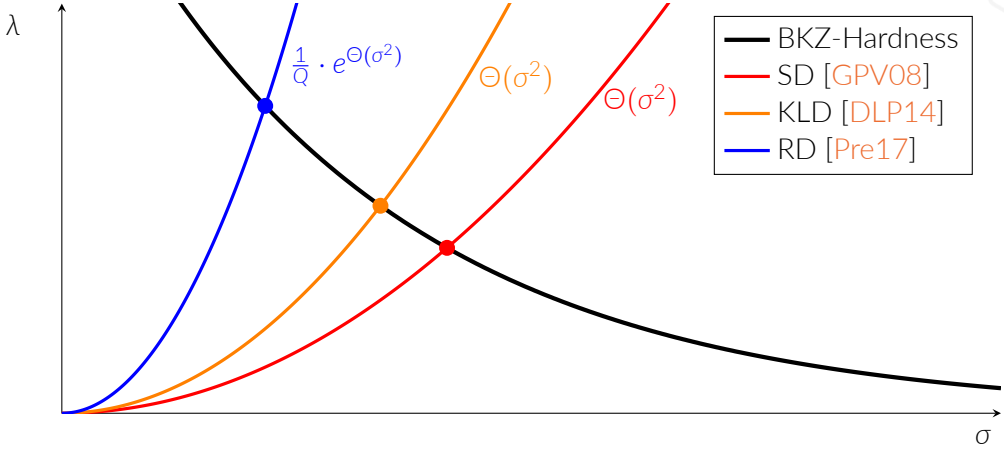


σ too large



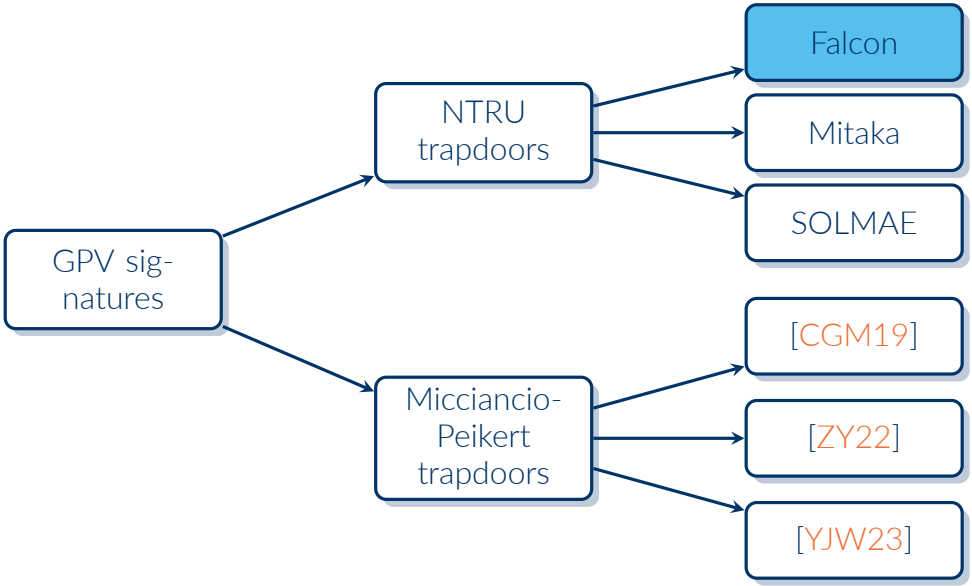
1. σ too small $\Rightarrow$ vulnerable to learning attacks [NR06, DN12]
2. σ too large $\Rightarrow$ suboptimal for cryptography

Standard deviation analysis



For $Q = 2^{64}$, we gain about 30 bits of security (compared to the SD).

Signature schemes in the GPV family



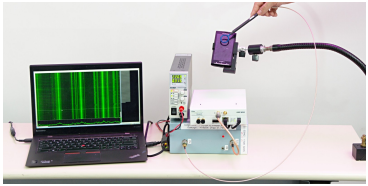
Falcon = GPV framework + NTRU trapdoors + Fast Fourier sampler + optimizations

Side-Channel Attacks

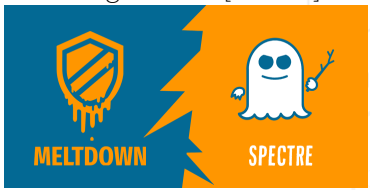
Side-channel attacks in cryptography

Power analysis attacks [KJJ99]

Electromagnetic attacks [Eck85]



Timing attacks [Koc96]



Acoustic attacks [AA04]

Visual attacks [NIC+23]

And more...

The return of the parallelepiped attacks

In Falcon, a signature **s** is distributed as a Gaussian.
The power consumption leaks information about the dot product $\langle \mathbf{s}, \mathbf{b}_0 \rangle$, or **s** itself.

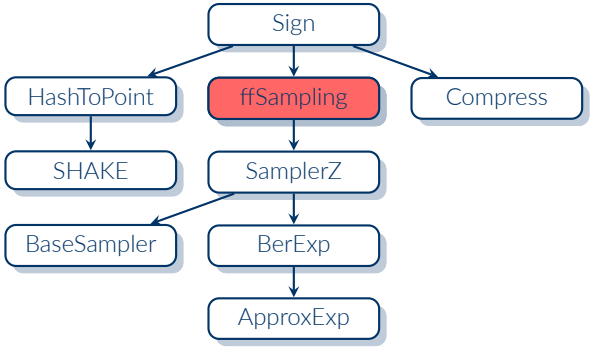
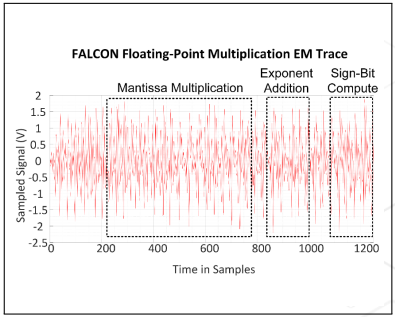


Figure 1: Flowchart of the signature



Learning **s** directly

[1] *FALCON Down: Breaking FALCON Post-Quantum Signature Scheme through Side-Channel Attacks* [KA21]

The return of the parallelepiped attacks

In Falcon, a signature **s** is distributed as a Gaussian.
The power consumption leaks information about the dot product $\langle \mathbf{s}, \mathbf{b}_0 \rangle$, or **s** itself.

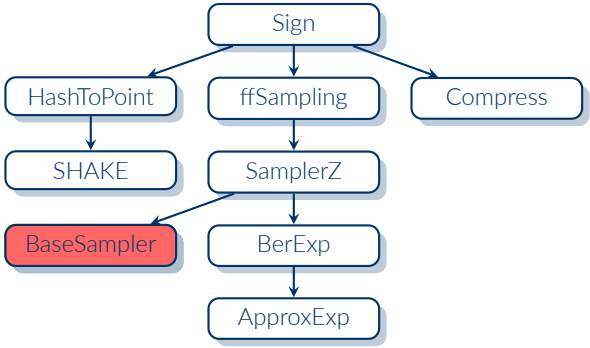
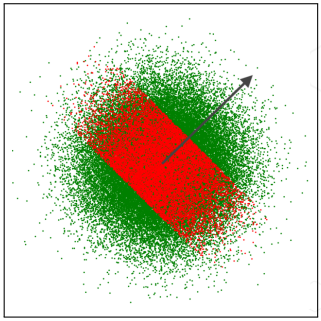


Figure 1: Flowchart of the signature



Filtering $\langle \mathbf{s}, \mathbf{b}_0 \rangle$ close to zero

[2] *The Hidden Parallelepiped Is Back Again: Power Analysis Attacks on Falcon* [GMRR22]

The return of the parallelepiped attacks

In Falcon, a signature **s** is distributed as a Gaussian.
The power consumption leaks information about the dot product $\langle \mathbf{s}, \mathbf{b}_0 \rangle$, or **s** itself.

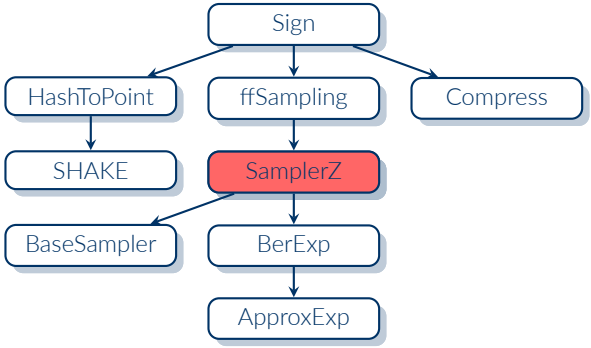
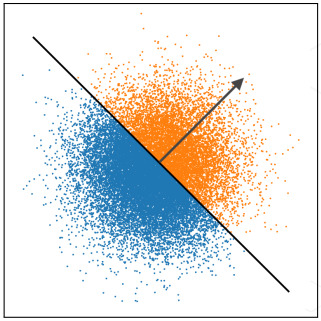


Figure 1: Flowchart of the signature



Filtering $\langle \mathbf{s}, \mathbf{b}_0 \rangle > 0$

[3] *Improved Power Analysis Attacks on Falcon* [ZLYW23]

Countermeasures

Against timing attacks: make signing isochronous (“cryptographic constant time”)

- BaseSampler reads a full table
- BerExp implements rejection sampling via polynomial approximation

The signing procedure is isochronous assuming that some basic FPA operations are.

Protection beyond timing attacks?

- [GMRR22, ZLYW23] propose countermeasures but they are ad hoc and only make their attacks more expensive to mount
- In general, the most robust countermeasure is masking
  - Masking Falcon is going to be very difficult
  - If masking is important, use Raccoon (`github.com/masksign/raccoon`)

When to Deploy

Pros and cons

Pros

- Compact public key and signature sizes
- Very fast verification
- Signing is also fast, but less than Dilithium

Cons

- Key generation and signing require FPA
  - Be mindful on devices with non-existent or variable-time FPA units
  - Say goodbye to masking
- Key generation and signing are complex to implement
- Key generation is slow-ish

Mapping Falcon properties to applications

PQ SHIELD

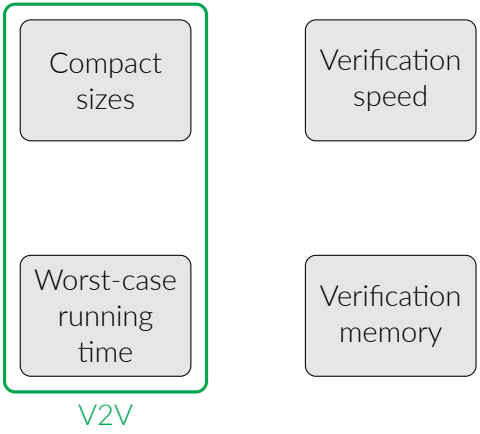
Compact sizes

Verification speed

Worst-case running time

Verification memory

Vehicle-to-vehicle (V2V) communications



Drive (Quantum) Safe! – Towards Post-Quantum Security for V2V Communications [BMTR22]

> " Only signature schemes whose explicit certificate can be sent in five or less fragments can be used in the *True Hybrid* design. [...] Falcon is the only viable scheme. "

TLS certificates

TLS

Compact sizes

Verification speed

Worst-case running time

Verification memory

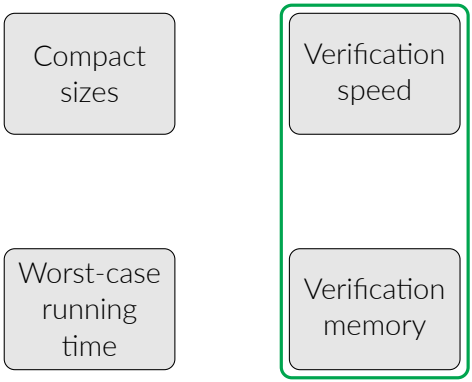
Post-Quantum Authentication in TLS 1.3: A Performance Study [SKD20]

> " The PQ algorithms with the best performance for time-sensitive applications are Dilithium and Falcon. "

NIST's pleasant post-quantum surprise [Wes22] recommends:

- → Falcon for offline signature
- → Dilithium for handshake

Verification on embedded devices

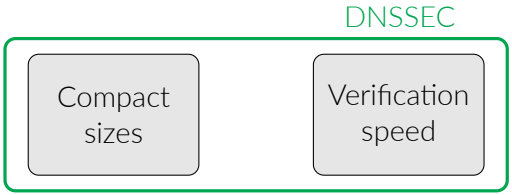


FPGA Energy Consumption of Post-Quantum Cryptography [BKG22]

> " For signature verification, Falcon provides the lowest energy consumption, highest throughput, and lowest transmission size [compared to Dilithium and SPHINCS+]. "

Verifying Post-Quantum Signatures in 8 kB of RAM [GHK+21]

> " On Cortex-M3, [Falcon's] overall memory footprint is about 6.5 kB. "



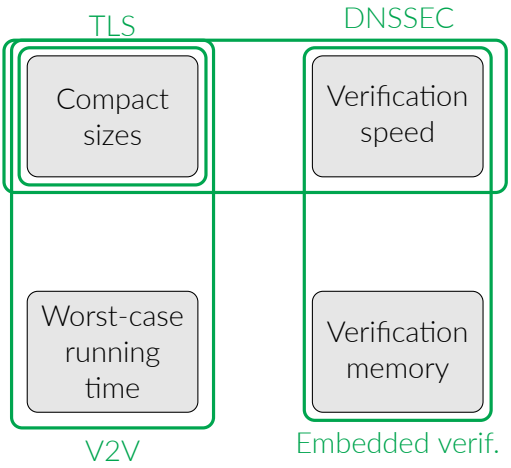
Retrofitting Post-Quantum Cryptography in Internet Protocols: A Case Study of DNSSEC [MdJvH+20]

> " [...] the performance of Falcon-512 is closest to the current algorithms and meets the requirements of DNSSEC. "

Post-Quantum Signatures in DNSSEC via Request-Based Fragmentation [GS22]

> " [...] Falcon-512 may be the most suitable option currently available to be standardized for DNSSEC. "

Summary



Suitable applications:

- V2V
- TLS certificates
- Verification on embedded devices
- DNSSEC
- ...

Questions?


Dmitri Asonov and Rakesh Agrawal.
Keyboard acoustic emanations.
In *2004 IEEE Symposium on Security and Privacy*, pages 3–11. IEEE Computer Society Press, May 2004.


Luke Beckwith, Jens-Peter Kaps, and Kris Gaj.
Fpga energy consumption of post-quantum cryptography.
In *Fourth PQC Standardization Conference*, 2022.
https://csrc.nist.gov/Events/2022/fourth-pqc-standardization-conference.


Nina Bindel, Sarah McCarthy, Geoff Twardokus, and Hanif Rahbari.
Drive (quantum) safe! — Towards post-quantum security for V2V communications.
Cryptology ePrint Archive, Report 2022/483, 2022.
https://eprint.iacr.org/2022/483.


Yilei Chen, Nicholas Genise, and Pratyay Mukherjee.
Approximate trapdoors for lattices and smaller hash-and-sign signatures.
In Steven D. Galbraith and Shiho Moriai, editors, *ASIACRYPT 2019*, *Part III*, volume 11923 of *LNCS*, pages 3–32. Springer, Heidelberg, December 2019.


Léo Ducas, Vadim Lyubashevsky, and Thomas Prest.
Efficient identity-based encryption over NTRU lattices.
In Palash Sarkar and Tetsu Iwata, editors, *ASIACRYPT 2014*, *Part II*, volume 8874 of *LNCS*, pages 22–41. Springer, Heidelberg, December 2014.

Léo Ducas and Phong Q. Nguyen.
Learning a zonotope and more: Cryptanalysis of NTRUSign countermeasures.
In Xiaoyun Wang and Kazue Sako, editors, *ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 433–450. Springer, Heidelberg, December 2012.

Wim Van Eck.
Electromagnetic radiation from video display units: An eavesdropping risk?
Computers & Security, 4:269–286, 1985.

Ruben Gonzalez, Andreas Hülsing, Matthias J. Kannwischer, Juliane Krämer, Tanja Lange, Marc Stöttinger, Elisabeth Waitz, Thom Wiggers, and Bo-Yin Yang.
Verifying post-quantum signatures in 8 kB of RAM.
In Jung Hee Cheon and Jean-Pierre Tillich, editors, *Post-Quantum Cryptography - 12th International Workshop, PQCrypto 2021*, pages 215–233. Springer, Heidelberg, 2021.

Morgane Guerreau, Ange Martinelli, Thomas Ricosset, and Mélissa Rossi.
The hidden parallelepiped is back again: Power analysis attacks on falcon.
IACR TCHES, 2022(3):141–164, 2022.

Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan.
Trapdoors for hard lattices and new cryptographic constructions.
In Richard E. Ladner and Cynthia Dwork, editors, *40th ACM STOC*, pages 197–206. ACM Press, May 2008.

Jason Goertzen and Douglas Stebila.

Post-quantum signatures in dnssec via request-based fragmentation, November 2022.


Emre Karabulut and Aydin Aysu.
FALCON down: Breaking FALCON post-quantum signature scheme through side-channel attacks.
In *58th ACM/IEEE Design Automation Conference, DAC 2021, San Francisco, CA, USA, December 5-9, 2021*, pages 691–696. IEEE, 2021.


Paul C. Kocher, Joshua Jaffe, and Benjamin Jun.
Differential power analysis.
In Michael J. Wiener, editor, *CRYPTO'99*, volume 1666 of *LNCS*, pages 388–397. Springer, Heidelberg, August 1999.


Paul C. Kocher.
Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems.
In Neal Koblitz, editor, *CRYPTO'96*, volume 1109 of *LNCS*, pages 104–113. Springer, Heidelberg, August 1996.


Patrick Karl, Jonas Schupp, Tim Fritzmann, and Georg Sigl.
Post-quantum signatures on risc-v with hardware acceleration.
ACM Trans. Embed. Comput. Syst., jan 2023.
Just Accepted.


Wai-Kong Lee, Raymond K. Zhao, Ron Steinfeld, Amin Sakzad, and Seong Oun Hwang.
High throughput lattice-based signatures on gpus: Comparing falcon and mitaka.
IACR Cryptol. ePrint Arch., page 399, 2023.


Moritz Müller, Jins de Jong, Maran van Heesch, Benno Overeinder, and Roland van Rijswijk-Deij.
Retrofitting post-quantum cryptography in internet protocols: A case study of dnssec.
volume 50, page 49–57, New York, NY, USA, oct 2020. Association for Computing Machinery.


Duc Tri Nguyen and Kris Gaj.
Fast falcon signature generation and verification using armv8 neon instructions.
In *AFRICACRYPT 2023*, 2023.
https://africacrypt2023.tn/accepted-papers/.


Ben Nassi, Etay Iluz, Or Cohen, Ofek Vayner, Dudi Nassi, Boris Zadov, and Yuval Elovici.
Video-based cryptanalysis: Extracting cryptographic keys from video footage of a device's power led.
Cryptology ePrint Archive, Paper 2023/923, 2023.
https://eprint.iacr.org/2023/923.


Phong Q. Nguyen and Oded Regev.
Learning a parallelepiped: Cryptanalysis of GGH and NTRU signatures.
In Serge Vaudenay, editor, *EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 271–288. Springer, Heidelberg, May / June 2006.


Thomas Pornin.
New efficient, constant-time implementations of falcon.
Cryptology ePrint Archive, Paper 2019/893, 2019.
https://eprint.iacr.org/2019/893.

Thomas Prest.
Sharper bounds in lattice-based cryptography using the Rényi divergence.
In Tsuyoshi Takagi and Thomas Peyrin, editors, *ASIACRYPT 2017, Part I*, volume 10624 of *LNCS*, pages 347–374. Springer, Heidelberg, December 2017.

Dimitrios Sikeridis, Panos Kampanakis, and Michael Devetsikiotis.
Post-quantum authentication in TLS 1.3: A performance study.
In *NDSS 2020*. The Internet Society, February 2020.

Bas Westerbaan.
Nist's pleasant post-quantum surprise.
The Cloudflare Blog, July 2022.
https://blog.cloudflare.com/nist-post-quantum-surprise/.

Yang Yu, Huiwen Jia, and Xiaoyun Wang.
Compact lattice gadget and its applications to hash-and-sign signatures.
Cryptology ePrint Archive, Paper 2023/729, 2023.
https://eprint.iacr.org/2023/729.

Shiduo Zhang, Xiuhan Lin, Yang Yu, and Weijia Wang.
Improved power analysis attacks on falcon.
Cryptology ePrint Archive, Paper 2023/224, 2023.
https://eprint.iacr.org/2023/224.


Shiduo Zhang and Yang Yu.
Towards a simpler lattice gadget toolkit.
In Goichiro Hanaoka, Junji Shikata, and Yohei Watanabe, editors, *PKC 2022, Part I*, volume 13177 of *LNCS*, pages 498–520. Springer, Heidelberg, March 2022.