NTRU Prime: round-3 updates

Daniel J. Bernstein
Billy Bob Brumley
Ming-Shing Chen
Chitchanok Chuengsatiansup
Tanja Lange
Adrian Marotzke
Bo-Yuan Peng
Nicola Tuveri
Christine van Vredendaal
Bo-Yin Yang

https://ntruprime.cr.yp.to

2021.06

Predicting post-quantum disasters

Deployed post-quantum system turns out to be breakable

$\Uparrow$

Nobody studied the relevant weakness in the attack surface

$\Uparrow$

Community didn't have enough time
to carefully review everything

$\Uparrow$

Careful security review was
much more complicated than necessary

$\Uparrow$

Design decisions didn't prioritize
minimizing the attack surface

Public attacks are still picking low-hanging fruit

Eurocrypt 2020 Bellare–Davis–Günther:
Instantaneous break of Round2 (a lattice submission with "provable security") via CCA hashing details.
CCA hashing details in NewHope "questionable".
CCA hashing details in ten submissions okay but "brittle".

Public attacks are still picking low-hanging fruit

Eurocrypt 2020 Bellare–Davis–Günther:
Instantaneous break of Round2 (a lattice submission with "provable security") via CCA hashing details.
CCA hashing details in NewHope "questionable".
CCA hashing details in ten submissions okay but "brittle".

Crypto 2020 Guo–Johansson–Nilsson:
Feasible attack, via timing leak from `memcmp`, against the "protected against timing and cache attacks" Frodo software.

Public attacks are still picking low-hanging fruit

Eurocrypt 2020 Bellare–Davis–Günther:
Instantaneous break of Round2 (a lattice submission with "provable security") via CCA hashing details.
CCA hashing details in NewHope "questionable".
CCA hashing details in ten submissions okay but "brittle".

Crypto 2020 Guo–Johansson–Nilsson:
Feasible attack, via timing leak from `memcmp`, against the "protected against timing and cache attacks" Frodo software.

Why were these attacks not published in 2017? 2018? 2019?
NISTPQC security reviewers are massively overloaded.
Focusing on round-3 candidates helps, but is it enough?

Lattice attacks keep getting better

2018 Laarhoven–Mariano saved "between a factor 20 to 40 in the time complexity for SVP".

2018 Bai–Stehlé–Wen introduced new variant of BKZ producing "bases of better quality" for "same cost" of SVP.

2018 Aono–Nguyen–Shen adapted "recent quantum tree algorithms" to enumeration.

2018 Anvers–Vercauteren–Verbauwhede showed that "an attacker can significantly reduce the security of (Ring/Module)-LWE/LWR based schemes that have a relatively high failure rate" and that for LAC-128 "the failure rate is 2^{48} times bigger than estimated".

Lattice attacks keep getting better, part 2

2018 Hamburg observed that the first published Round5 design had disastrously high failure rate, 2^{-55}.

2019 Pellet-Mary–Hanrot–Stehlé broke through the previously claimed half-exponential approximation-factor barrier for number-theoretic attacks against Ideal-SVP.

2019 Guo–Johansson–Yang presented faster attacks against some systems that use error correction to (try to) reduce decryption failures. Violated security claims of LAC.

2020 Bellare–Davis–Günther broke Round2. (See above.)

2020 Dachman-Soled–Ducas–Gong–Rossi presented slightly faster attacks against the constant-sum secrets used in three lattice submissions: LAC, NTRU, Round5.

Lattice attacks keep getting better, part 3

2020 Doulgerakis–Laarhoven–de Weger presented "faster [sieving] methods" for SVP.

2020 Albrecht–Bai–Fouque–Kirchner–Stehlé–Wen reduced the exponent of enumeration from $\approx\mathbf{0.187}\beta\log_2\beta$ to $\approx\mathbf{0.125}\beta\log_2\beta$. Combined with 2018 Aono–Nguyen–Shen, reduces post-quantum security levels of lattice-based systems.

2020 Bernard–Roux-Langlois improved the algorithm of 2019 Pellet-Mary–Hanrot–Stehlé; showed experimentally that in small dimensions the improved algorithm reaches much better approximation factors. How well does this scale to larger dimensions?

So what do we do?

Sensible reaction to drumbeat of advances in lattice attacks:
Lattices are dangerous! Avoid them!
Why were we considering lattices in the first place?

So what do we do?

Sensible reaction to drumbeat of advances in lattice attacks:
Lattices are dangerous! Avoid them!
Why were we considering lattices in the first place?

1. Common answer: "provable security". But this didn't stop any of the post-2017 advances in attacks, didn't save the broken systems; won't stop further advances.

So what do we do?

Sensible reaction to drumbeat of advances in lattice attacks:
Lattices are dangerous! Avoid them!
Why were we considering lattices in the first place?

1. Common answer: "provable security". But this didn't stop any of the post-2017 advances in attacks, didn't save the broken systems; won't stop further advances.

2. Much better argument specifically for small lattices:
Maybe the application requires a small lattice system—
Frodo is too big, SIKE is too slow, etc.

NTRU Prime's core advantage

Subject to the small-lattice requirement,
NTRU Prime is the only submission systematically designed to **eliminate unnecessary complications in security review:** eliminate decryption failures, eliminate cyclotomics, etc.

NTRU Prime's core advantage

Subject to the small-lattice requirement,
NTRU Prime is the only submission systematically designed to **eliminate unnecessary complications in security review:** eliminate decryption failures, eliminate cyclotomics, etc.

Does this work? Yes: every improvement in NTRU Prime attacks has been a general improvement against small lattices, while various small-lattice submissions have suffered from classes of attacks that NTRU Prime had already eliminated.

Example: decryption failures

Many attack advances. Some security claims broken.
More and more pages of increasingly complicated analysis.

2021 D'Anvers–Batsleer: "We first improve the state-of-the-art multitarget decryption failure attack using a levelled approach", point out "three inaccuracies in the directional failure boosting calculation for the simplified scheme of [11]", show that "this traditional approach of calculating the directional failure boosting cost is not directly applicable to practical schemes such as Kyber and Saber due to compression of the ciphertexts", etc.

Example: decryption failures

Many attack advances. Some security claims broken.
More and more pages of increasingly complicated analysis.

2021 D'Anvers–Batsleer: "We first improve the state-of-the-art multitarget decryption failure attack using a levelled approach", point out "three inaccuracies in the directional failure boosting calculation for the simplified scheme of [11]", show that "this traditional approach of calculating the directional failure boosting cost is not directly applicable to practical schemes such as Kyber and Saber due to compression of the ciphertexts", etc.

NTRU Prime, 2014: "prefer to avoid the mess of figuring out whether an attacker can trigger decryption failures."

Proactive, not reactive

Proactive

Attack tool is identified

↓

Design systems to eliminate the tool

Reactive

Attack tool is identified

↓

First attacks are developed

↓

Better attacks are developed

↓

Devastating attacks are developed

↓

Design systems to eliminate the attacks

Example: cyclotomics

2013.07: Cyclotomic lattices give subfields, automorphisms, etc. to attacker. Conclusion: “should switch” from cyclotomics to “random prime-degree extensions with big Galois groups”.

Example: cyclotomics

2013.07: Cyclotomic lattices give subfields, automorphisms, etc. to attacker. Conclusion: "should switch" from cyclotomics to "random prime-degree extensions with big Galois groups".

2014.02: First complete NTRU Prime cryptosystem, with maximum-size Galois groups and no decryption failures. Also introduced subfield-logarithm attack—subexponential time against some extreme cases of STOC 2009 Gentry FHE.

All of this was *before* poly-time break of cyclotomics for the FHE system, and before the NISTPQC decryption-failure mess.

Example: cyclotomics

2013.07: Cyclotomic lattices give subfields, automorphisms, etc. to attacker. Conclusion: "should switch" from cyclotomics to "random prime-degree extensions with big Galois groups".

2014.02: First complete NTRU Prime cryptosystem, with maximum-size Galois groups and no decryption failures. Also introduced subfield-logarithm attack—subexponential time against some extreme cases of STOC 2009 Gentry FHE.

All of this was *before* poly-time break of cyclotomics for the FHE system, and before the NISTPQC decryption-failure mess.

(Very similar recommendations to avoid subfields and automorphisms in discrete-log cryptography were published years before those features turned into devastating attacks.)

Example: cyclotomics

2013.07: Cyclotomic lattices give subfields, automorphisms, etc. to attacker. Conclusion: "should switch" from cyclotomics to "random prime-degree extensions with big Galois groups".

2014.02: First complete NTRU Prime cryptosystem, with maximum-size Galois groups and no decryption failures. Also introduced subfield-logarithm attack—subexponential time against some extreme cases of STOC 2009 Gentry FHE.

All of this was *before* poly-time break of cyclotomics for the FHE system, and before the NISTPQC decryption-failure mess.

(Very similar recommendations to avoid subfields and automorphisms in discrete-log cryptography were published years before those features turned into devastating attacks.)

NIST's statements get this history backwards, hiding the cryptographer's ability to proactively protect against risks.

Four plausible scenarios for the future of attacks

Attacks published so far:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	ok	ok

Four plausible scenarios for the future of attacks

Attacks published so far:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	ok	ok

If cyclotomics are weak:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	**broken**	ok

Four plausible scenarios for the future of attacks

Attacks published so far:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	ok	ok

If cyclotomics are weak:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	**broken**	ok

If Gentry's system is weak:

	Cyclo	Prime
Gentry	**broken**	**broken**
KEM	ok	ok

Four plausible scenarios for the future of attacks

Attacks published so far:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	ok	ok

If cyclotomics are weak:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	**broken**	ok

If Gentry's system is weak:

	Cyclo	Prime
Gentry	**broken**	**broken**
KEM	ok	ok

If lattices are weak:

	Cyclo	Prime
Gentry	**broken**	**broken**
KEM	**broken**	**broken**

Four plausible scenarios for the future of attacks

Attacks published so far:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	ok	ok

If cyclotomics are weak:

	Cyclo	Prime
Gentry	**broken**	ok
KEM	**broken**	ok

If Gentry's system is weak:

	Cyclo	Prime
Gentry	**broken**	**broken**
KEM	ok	ok

If lattices are weak:

	Cyclo	Prime
Gentry	**broken**	**broken**
KEM	**broken**	**broken**

Four further scenarios—split between showing security advantages of NTRU Prime and showing security disadvantages of NTRU Prime—are logically possible but harder to explain as developments from the current situation.

Is it still small? Yes—often even higher security!

Example: Say an application requires $\leq$1024 bytes for each public key and each ciphertext.

Is it still small? Yes—often even higher security!

Example: Say an application requires $\leq$1024 bytes
for each public key and each ciphertext.
Maximum achievable pre-quantum $\log_2$ Core-SVP level
across all of the parameter sets proposed for standardization:

- 106 with NTRU (`ntruhps2048509`).
- 112 (claimed 118?) with (round-3) Kyber (`kyber512`).
- 118 with SABER (`lightsaber`).
- **129** with Streamlined NTRU Prime (`sntrup653`).

Is it still small? Yes—often even higher security!

Example: Say an application requires $\leq$1024 bytes for each public key and each ciphertext.
Maximum achievable pre-quantum $\log_2$ Core-SVP level across all of the parameter sets proposed for standardization:

- 106 with NTRU (`ntruhps2048509`).
- 112 (claimed 118?) with (round-3) Kyber (`kyber512`).
- 118 with SABER (`lightsaber`).
- **129** with Streamlined NTRU Prime (`sntrup653`).

Vary 1024 $\Rightarrow$ sometimes (e.g.) Kyber has higher Core-SVP.
The big picture is that each candidate is competitive.
See https://cr.yp.to/papers.html#categories.

Another aspect of sizes: fitting into hardware

CARDIS 2020: **Complete** constant-time `sntrup761` using 1841 FPGA slices with 14 BRAMs, 19 DSPs.

(Xilinx Zynq Ultrascale+: Artix-7, plus ARM not used here.)

Update: **Complete** constant-time `sntrup761` using 1367 FPGA slices with 11.5 BRAMs, 19 DSPs.

Runs at 271.6 MHz. Cycle counts:
1289959 keygen, 119250 enc, 260307 dec.

`https://github.com/AdrianMarotzke/SNTRUP`

With somewhat more area, can achieve far fewer cycles;
but TCO analysis suggests that small area is more important.

NTRU Prime speeds are competitive

Almost all KEMs reported in pqm4 (as of 2021.06.02) have slower enc and slower dec than sntrup761.

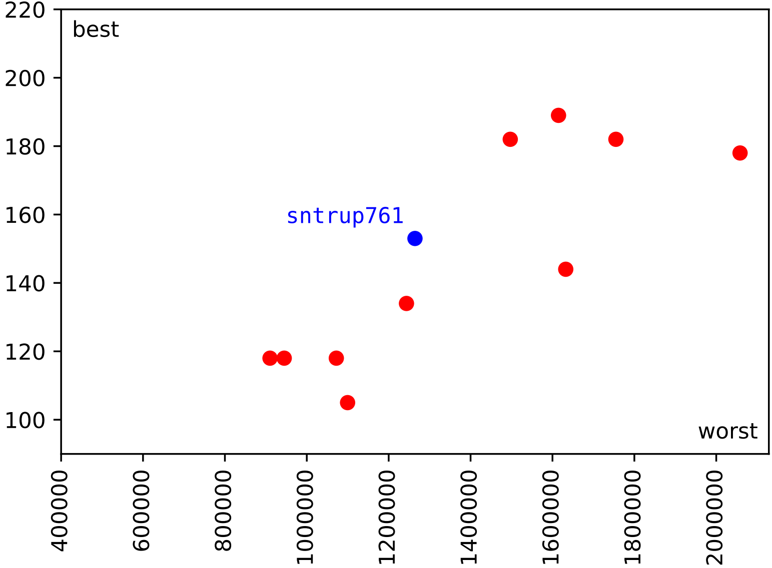
NTRU Prime speeds are competitive

Almost all KEMs reported in `pqm4` (as of 2021.06.02)
have slower enc and slower dec than `sntrup761`.

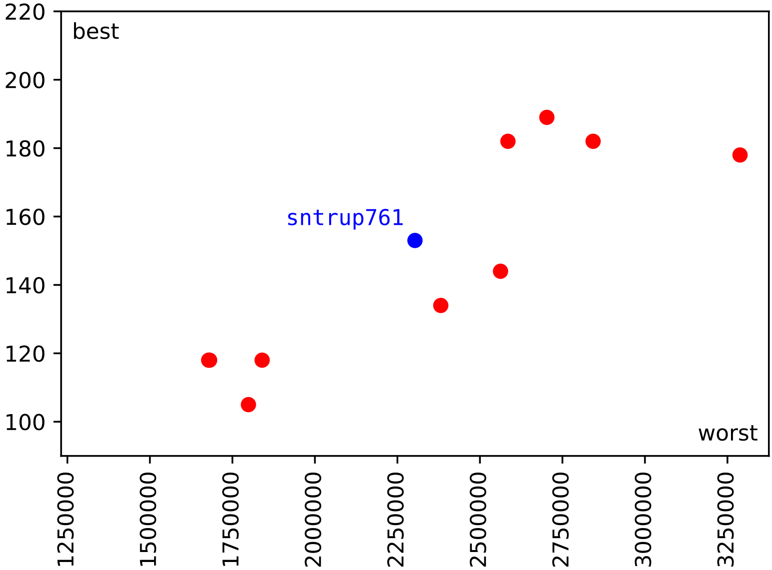
Pre-quantum $\log_2$ Core-SVP levels for the exceptions:

- 106 for `ntruhps2048509`: 563499 enc, 536107 dec.
- 112 (118?) for `kyber90s512`: 449428 enc, 460732 dec.
- 112 (118?) for `kyber512`: 555947 enc, 516170 dec.
- 118 for `lightsaber`: 484733 enc, 460133 dec.
- 136 for `ntruhrss701`: 375974 enc, 867459 dec.
- **153** for `sntrup761`: 698943 enc, 565268 dec.

Security (vertical) vs. enc + dec time (horizontal)



Security vs. enc time + dec time + 1000 · ctbytes



Stability: another win for security reviewers

NTRU Prime has **an unchanged family of trapdoor functions throughout round 1, round 2, and round 3.** See round-3 submission for analysis of how modules, errors, etc. would complicate security review.

CCA conversion included various hashing safeguards in round 1. Added further defenses in round 2. Unchanged in round 3. $\Rightarrow$ NTRU Prime is **fully compatible between round 2 and round 3, when users choose the same parameters**.

Have always recommended the same parameter set: dimension $p = 761$, modulus $q = 4591$.

Many fully specified+implemented parameter sets

Dimension:	653	761	857	953	1013	1277
Pre-quantum $\log_2$ Core-SVP:	129	153	175	196	209	270
Post-quantum $\log_2$ Core-SVP:	117	139	159	178	190	245

See submission for other metrics, analysis of Core-SVP flaws.

Given likelihood of further advances in
lattice attacks affecting all small-lattice submissions,
NTRU Prime will not add dimensions below 653.
761 is recommended for an extra security margin.

Round-1 NTRU Prime specified *only* 761. This type of focus simplifies implementations, simplifies security analysis. However, NIST keeps asking for more parameter sets.

Quotient NTRU vs. Product NTRU

Two fully supported options at each size—
picking just one would need more transparency from NIST:

- Streamlined NTRU Prime (example of Quotient NTRU).
  As in NTRU submission, security risks from homogeneity.
  No known patent threats. NTRU patent expired 2017.
- NTRU LPRime (example of Product NTRU).
  As in Kyber+SABER, security risks from extra samples,
  and from looseness of known QROM IND-CCA2 proofs.
  Threatened by same patents as Kyber+SABER;
  see https://ntruprime.cr.yp.to/faq.html.

Extensive sharing of code and analysis across these options.

Does Quotient NTRU have slow keygen?

Haswell cycles/key	enc	dec	`sntrup761` report
$>$6000000	59456	97684	round-1 submission

Does Quotient NTRU have slow keygen?

Haswell cycles/key	enc	dec	sntrup761 report
>6000000	59456	97684	round-1 submission
946772	55252	70464	round-2 update talk

Does Quotient NTRU have slow keygen?

Haswell cycles/key	enc	dec	`sntrup761` report
>6000000	59456	97684	round-1 submission
946772	55252	70464	round-2 update talk
156317	46914	56241	now

Does Quotient NTRU have slow keygen?

Haswell cycles/key	enc	dec	`sntrup761` report
>6000000	59456	97684	round-1 submission
946772	55252	70464	round-2 update talk
156317	46914	56241	now

Haswell cycles/key	enc	dec	`ntrulpr761` report
47396	77280	95316	round-2 update talk
42515	69103	82071	now

Does Quotient NTRU have slow keygen?

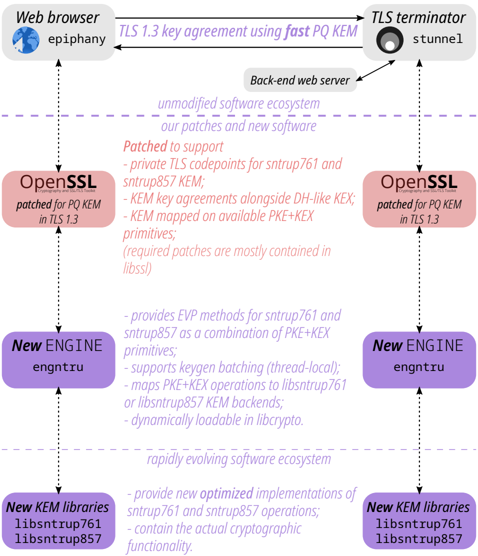
Haswell cycles/key	enc	dec	`sntrup761` report
>6000000	59456	97684	round-1 submission
946772	55252	70464	round-2 update talk
156317	46914	56241	now

Haswell cycles/key	enc	dec	`ntrulpr761` report
47396	77280	95316	round-2 update talk
42515	69103	82071	now

These times are not a bottleneck in any known application.

Web-browsing demo using `sntrup761` with fast keygen:
`https://opensslntru.cr.yp.to`

OpenSSLNTRU software stack



What exactly does “small” mean?

If the larger NTRU Prime parameters still count as “small”, shouldn’t we also be considering Frodo?
Maybe small lattices are weak but Frodo survives!

What exactly does “small” mean?

If the larger NTRU Prime parameters still count as “small”, shouldn’t we also be considering Frodo?
Maybe small lattices are weak but Frodo survives!

But anyone who can afford Frodo-m can afford `sntrup`-N or `ntrulpr`-N with much higher security against known attacks.
Would an unknown attack be able to close this gap? Unclear.

	`frodokem640`	`ntrulpr1277`
key bytes	9616	1847
ciphertext bytes	9720	1975
keygen cycles (Haswell)	1490605	77092
enc cycles	1922241	121397
dec cycles	1849960	144582
Core-SVP	150	270

Community confidence in cyclotomic lattices?

Many round-1 lattice KEMs were NewHope variants,
so they used cyclotomics by default.

Community confidence in cyclotomic lattices?

Many round-1 lattice KEMs were NewHope variants,
so they used cyclotomics by default.

But 41% of the round-1 lattice submissions

- provided options that do *not* use cyclotomics;
- in most cases paid heavily in performance for this;
- in most cases expressed concerns regarding security;
- in most cases provided *no* cyclotomic options.

Community confidence in cyclotomic lattices?

Many round-1 lattice KEMs were NewHope variants,
so they used cyclotomics by default.

But 41% of the round-1 lattice submissions

- provided options that do *not* use cyclotomics;
- in most cases paid heavily in performance for this;
- in most cases expressed concerns regarding security;
- in most cases provided *no* cyclotomic options.

2020 NIST report refers to "NIST's confidence in cyclotomic structures". Where does this confidence come from? 3 years?

Is NIST actually confident in Kyber?

NIST says Frodo is "a conservative backup in the case of new cryptanalytic results targeting structured lattices being discovered in the third round."

Why just a backup? Why not proactively standardize Frodo? Is this because NIST claims the risks are negligible?

No: NIST says performance. "NIST's first priority for standardization is a KEM that would have acceptable performance in widely used applications overall."

If NIST were actually confident in Kyber then wouldn't it have said "We don't see a need for Frodo" and eliminated it?

What NIST was thinking

NIST's AES reports had comparison charts, surveys of attacks, etc. NIST's reports for NISTPQC are much less detailed.

What NIST was thinking

NIST's AES reports had comparison charts, surveys of attacks, etc. NIST's reports for NISTPQC are much less detailed.

Fortunately, a NIST employee gave a talk on 2020.08.28 explaining "NIST's decision process for Round 3".

Unfortunately, no public announcement of talk.
NIST later described the talk as having been
"given to the University of Maryland Crypto Reading Group".

What NIST was thinking + didn't want to tell us

NIST's AES reports had comparison charts, surveys of attacks, etc. NIST's reports for NISTPQC are much less detailed.

Fortunately, a NIST employee gave a talk on 2020.08.28 explaining "NIST's decision process for Round 3".

Unfortunately, no public announcement of talk.
NIST later described the talk as having been
"given to the University of Maryland Crypto Reading Group".

In fact, NIST had invited *some* round-3 lattice submitters to attend the talk and the subsequent Q&A session: Dilithium ($2\times$), Falcon ($2\times$), Frodo, Kyber ($3\times$), NTRU ($2\times$), SABER.
NIST also invited a SIKE submitter and some others.

Talk slides were posted in response to a FOIA request.

What NIST was thinking: the details

NIST's slides say: "among" the "remaining, structured" lattice KEMs, "our assessment was that cyclotomics (esp power-of-2 cyclotomics) are the clear 'community standard' ".

What NIST was thinking: the details

NIST's slides say: "among" the "remaining, structured" lattice KEMs, "our assessment was that cyclotomics (esp power-of-2 cyclotomics) are the clear 'community standard' ".

It's clear that if we ignore Titanium etc. as not "remaining", and ignore Frodo etc. as not being "structured" lattice KEMs, then 3 of the 4 candidates use cyclotomics.

What NIST was thinking: the details

NIST's slides say: "among" the "remaining, structured" lattice KEMs, "our assessment was that cyclotomics (esp power-of-2 cyclotomics) are the clear 'community standard' ".

It's clear that if we ignore Titanium etc. as not "remaining", and ignore Frodo etc. as not being "structured" lattice KEMs, then 3 of the 4 candidates use cyclotomics.

Could just as easily exclude any other targeted submission by selecting a distinguishing feature of that submission and claiming that the feature is not the "community standard".

What NIST was thinking: the details

NIST's slides say: "among" the "remaining, structured" lattice KEMs, "our assessment was that cyclotomics (esp power-of-2 cyclotomics) are the clear 'community standard' ".

It's clear that if we ignore Titanium etc. as not "remaining", and ignore Frodo etc. as not being "structured" lattice KEMs, then 3 of the 4 candidates use cyclotomics.

Could just as easily exclude any other targeted submission by selecting a distinguishing feature of that submission and claiming that the feature is not the "community standard".

Should instead be proactively minimizing risks.

Ongoing work on NTRU Prime

If public analysis doesn't completely break small lattices:
We expect continuing interest from SDOs and users who
(1) think they need small lattices for performance but
(2) still don't want to incur unnecessary security risks.

We're working on important preparations for real-world usage:
porting, formal verification, masking, etc.

Ongoing work on NTRU Prime

If public analysis doesn't completely break small lattices:
We expect continuing interest from SDOs and users who
(1) think they need small lattices for performance but
(2) still don't want to incur unnecessary security risks.

We're working on important preparations for real-world usage: porting, formal verification, masking, etc.

What about NIST? Is that one of these SDOs?

— NIST is disregarding the known cyclotomic attacks; hoping that STOC 2009 Gentry FHE isn't the canary in the coal mine; and standardizing cyclotomic KEMs by default.

Ongoing work on NTRU Prime

If public analysis doesn't completely break small lattices:
We expect continuing interest from SDOs and users who
(1) think they need small lattices for performance but
(2) still don't want to incur unnecessary security risks.

We're working on important preparations for real-world usage: porting, formal verification, masking, etc.

What about NIST? Is that one of these SDOs?

— NIST is disregarding the known cyclotomic attacks; hoping that STOC 2009 Gentry FHE isn't the canary in the coal mine; and standardizing cyclotomic KEMs by default.

However, if big enough advances in cyclotomic attacks appear this year, NIST will put cyclotomics on hold, see whether NTRU Prime survives round 4, and consider selecting it.