From: (b) (6)
To: Perlner, Ray A. (Fed)
Cc: Apon, Daniel C. (Fed); Dang, Quynh H. (Fed); Moody, Dustin (Fed); internal-pqc; Dang, Thinh H. (Fed)
Subject: Re: Kyber"s response discussion tomorrow ?
Date: Friday, June 5, 2020 12:07:29 PM

I think that I agree with everything there. To me I think that we need to raise the issue a bit more directly with Dilithium and Falcon. It is definitely too generous to simply not have the same requirements for lattice signatures that we have for literally every other scheme, whether KEM or sig.

On Fri, Jun 5, 2020 at 11:47 AM Perlner, Ray A. (Fed) <ray.perlner@nist.gov> wrote:

> I suppose I should say more than simply "today's technology is orders of magnitude closer to making 2^89 bits of memory feasible than 2^143 bit operations of computation." I think it boils down to what is the point of having security levels above what's technologically feasible. For me, I see it as primarily a hedge against both improvements in technology and improvements in cryptanalysis.
>
> With regard to technological improvement, it seems to me that the cost per bit of memory is currently dropping faster than the cost per bit operation of computation, and theoretically, it seems like this should be the case since the cost of computation is constrained by the Landauer limit, which is only something like 4 orders of magnitude below what's currently achievable, while the constraints for how good a memory can be, seem, to me at least, a bit squishier. There is the loophole of reversible computing, but that requires slower computation, hence more parallelism, i.e. cheaper memory, anyway. Thus, even if it seems like the memory and compute requirements of an attack seem comparable with today's technology, I'd be inclined to consider the computation to be the limiting cost for future technology.
>
> With regard to algorithmic improvement, I can think of a number of advances where memory requirements have dramatically dropped with modest to nonexistent effect on overall computational cost (e.g. van Oorschot-Wiener, and block Wiedemann – especially when the matrix being inverted can be cheaply regenerated on the fly.) That may be selection bias, though, if the computational cost of an attack drops dramatically, it's not cryptographically interesting anymore. I could be convinced this is anything from a wash to a slight point against putting too much reliance on memory cost of attacks to provide security.
>
> Now, with Kyber, the raw bit operation cost they're estimating is pretty close to what we're asking for (off by 2^7 by the less generous method) so I'm inclined to think things like the cost of random access queries really can cover the cost even in a fairly speculative model of how good a memory can be. I'm less sure about Dilithium and Falcon which have to close a gap of something like 2^17 or 2^14. NTRU is between these extremes.

There's also the fairness issue of treating things that have 20+ bits more security margin as exactly the same, and the issue of what to do about the higher security levels. Differences between higher security levels are of course less likely to be practically relevant, but the difference in coreSVP is much larger in terms of what the candidates are claiming is level 5 as opposed to what they are claiming is level 1. (e.g. NTRUprime's claimed category 5 parameters are below Kyber's claimed category 3 parameters.) We definitely should ask for more consistency with the higher security levels or at least publicly note that we will take the gap in claims seriously when making comparisons between higher level parameters.

From: Perlner, Ray A. (Fed) <ray.perlner@nist.gov>
Sent: Friday, June 5, 2020 11:11 AM
To: Apon, Daniel C. (Fed) <daniel.apon@nist.gov>; Daniel Smith (b) (6) ; Dang, Quynh H. (Fed) <quynh.dang@nist.gov>
Cc: Moody, Dustin (Fed) <dustin.moody@nist.gov>; internal-pqc <internal-pqc@nist.gov>; Dang, Thinh H. (Fed) <thinh.dang@nist.gov>
Subject: RE: Kyber's response discussion tomorrow ?

Just as an aside. I'm supposed to be impressed by "**A planar sheet of terabyte micro-SD cards the size of New York City (all five boroughs, 800 km^2 ~ 2^49.5 mm^2) would hold 2^89 bits.**"?

Yes. It's a big number, but the computational complexity we're asking for even in category 1 is HUGE. I did some calculations. Supposing you didn't just blanket new York city with solar panels operating at 20% efficiency, but the entire continent of north America. That would get us about 2^51 W. I did a quick Google for a high-end bitcoin miner, and found Antminer S17+ advertising 73TH/s at 2920W. Assuming 1 Hash is a double SHA2 operation, i.e. 2^19 bit operations, this corresponds to 2^53.5 bit operations per second per watt. Putting this together with our power budget, that's 2^104.5 bit operations per second. There are 2^25 seconds in a year. Thus category 1's classical security requirement of 2^143 bit operations comes out to the computational capacity of an array of top end bitcoin miners powered by tiling North America with solar panels running for 10000 years.

From: Apon, Daniel C. (Fed) <daniel.apon@nist.gov>
Sent: Friday, June 5, 2020 3:41 AM

To: Perlner, Ray A. (Fed) <ray.perlner@nist.gov>; Daniel Smith (b) (6) ; Dang, Quynh H. (Fed) <quynh.dang@nist.gov>
Cc: Moody, Dustin (Fed) <dustin.moody@nist.gov>; internal-pqc <internal-pqc@nist.gov>; Dang, Thinh H. (Fed) <thinh.dang@nist.gov>
Subject: RE: Kyber's response discussion tomorrow ?

I felt that the key point in the Kyber Team's response was

"We agree that … 141 [is] smaller than 143, but at the moment we do not consider this to be a sufficient reason to modify the Kyber-512 parameter set.
…
The additional memory requirement of this attack strongly suggests that Kyber-512 is more secure than AES-128 in any realistic cost model.
…
A planar sheet of terabyte micro-SD cards the size of New York City (all five boroughs, 800 km^2 ~ 2^49.5 mm^2) would hold 2^89 bits.
"

I still feel we should do our own internal analysis at the start of the 3rd Round.

I'm utterly opposed to letting DJB's eleventh-hour protestations influence absolutely anything whatsoever.

--Daniel

From: Perlner, Ray A. (Fed) <ray.perlner@nist.gov>
Sent: Thursday, June 4, 2020 3:02 PM
To: Daniel Smith (b) (6) ; Dang, Quynh H. (Fed) <quynh.dang@nist.gov>
Cc: Moody, Dustin (Fed) <dustin.moody@nist.gov>; internal-pqc <internal-pqc@nist.gov>; Dang, Thinh H. (Fed) <thinh.dang@nist.gov>
Subject: RE: Kyber's response discussion tomorrow ?

Here are my current thoughts on the matter:

I am open to the idea of using a more realistic model of computation than the basic gate model. However, a lot of the ideas in I've seen in the literature seem too pessimistic (as in they reckon attacks as being harder than they should be. – at least in the long term)

DJB's favored model, for example, assumes the computation must be implemented by only nearest neighbor interactions in a 2 dimensional grid. This has some justification, in that

trying to violate these assumptions clearly costs more than the basic gate model assumes, but

1. Today's Supercomputers generally use a meaningfully 3 dimensional arrangement of processors (although the processors themselves are 2 dimensional)
2. Long distance connections needing high performance are implemented by fiber optic cables, and sending a bit through a kilometer of fiber optic cable, while more expensive than sending the bit across a single AND gate, clearly costs less than sending it through a kilometer of densely packed AND gates (which is how DJB's favored model would treat it.)

NTRU's "local" model seems in practice to be even more extreme, simply ignoring any algorithm that hasn't explicitly been implemented locally

Hard limits on the total memory size have also been proposed. I think the smallest numbers I could really convince myself were commensurate with an adversary actually capable of threatening the appropriate security level were 2^100 for levels 1 and 2, 2^150 for levels 3 and 4 , and 2^180 for level 5.

One could perhaps adjust the RAM model to cost random access queries to a memory of size N at N^(1/3) in terms of depth and (log(N))^2 in terms of gate count and require all other gates to be local. (I think I might actually be ok with that, keeping in mind that if the whole thing can be implemented locally, you don't need to make RAM queries, no matter how large the computation is.)

The other worry though is that things like memory cost are much more susceptible to being optimized away by incremental improvements, which the first iteration of a new attack rarely includes. But there are a lot of smart lattice people, so maybe I can be convinced they've thought about this stuff enough that there is no room for further improvement. I'm not convinced yet, though.

Ray

From: Daniel Smith (b) (6)
Sent: Thursday, June 4, 2020 2:23 PM
To: Dang, Quynh H. (Fed) <quynh.dang@nist.gov>
Cc: Moody, Dustin (Fed) <dustin.moody@nist.gov>; internal-pqc <internal-pqc@nist.gov>; Dang, Thinh H. (Fed) <thinh.dang@nist.gov>
Subject: Re: Kyber's response discussion tomorrow ?

Hmmm...

They are calling us out explicitly to offer our position on this. It is a muddy issue in my mind.

I have a bit of a problem with saying, "We are secure because of other stuff that we can't measure really well." For other areas we have been requiring them to ignore memory costs even when that makes a difference for them.

A clear example comes to mind: GeMSS. For GeMSS they had a quite exhaustive analysis of known techniques applied to GeMSS. They quite conservatively used analyses and coefficients that are unrealistic even with zero cost of memory and memory access (which is why confusingly they chose to report some of the numbers as lower than the security bounds when actually they should be fine). When you consider the hidden polynomial factors or actual coefficients, the least costly attack (and the one they are basing the parameters on) is the direct algebraic attack. They are being super conservative and choosing a linear algebra exponent of 2 for dense linear algebra (I think that we can't use sparse techniques here because of the number of solutions (or the density after fixing variables)), but if we take memory into account, then the complexity is altogether different. If our metric is New York City, then this scheme should benefit fairly significantly.

On a historical note, Ray and I argued fairly extensively about this memory issue when we were drafting the CFP. I recall having discussions about the physical feasibility of converting Jupiter into atomic scale memory that violates causality with the speed of its access (sending replies and being set to different values before being asked to) leading up to the release of this document. The issue as I recall was allowing the community to address some complexity issues that had not been pinned down yet at the time and for the community to come to a consensus on how to address these things. Still, we need to have some standard metric for comparisons between schemes.

I think that it is entirely reasonable to address memory and memory access in a cost model. A problem occurs when we lack justification and when we lack consistency in how we apply restrictions in these analyses. Ray and I were arguing on the level of Jupiter and breaking the laws of physics, whereas Kyber is arguing on the level of the 5 boroughs.

I would be open to allowing teams to specify their cost model addressing memory (in communication with us and with clear justification and theoretical support), and to adjust parameters accordingly. This would need to take place extremely quickly, though, to not make analysis placed on a moving target.

The easiest way to handle the situation is exactly the opposite, though. That is to let the teams do what they are doing and then judge them by our own metrics. The downside of this approach is that there is plenty of room for bias and plenty of reason for skepticism in our choices if any parts of our community think that we are cutting corners unreasonably.

If we chose to allow memory access cost as part of the complexity analysis, there will be consequences. We may have to communicate with each team explicitly, but I think we should make it clear (if we go that route) that they should analyze the memory concerns with strong justification for **minimal** cost models that they can then incorporate. We also need to assess the feasibility of these models and the appropriateness of the bounds they suggest.

I think that we have plenty to talk about, but we'll follow your lead, Dustin.

Cheers,

Daniel

On Thu, Jun 4, 2020 at 1:47 PM Dang, Quynh H. (Fed) <quynh.dang@nist.gov> wrote:

> I think so. If more people think that a talk tomorrow would be good, then I would ask you to consider that.
>
> ---
>
> **From:** Moody, Dustin (Fed) <dustin.moody@nist.gov>
> **Sent:** Thursday, June 4, 2020 1:41 PM
> **To:** Dang, Quynh H. (Fed) <quynh.dang@nist.gov>
> **Cc:** internal-pqc <internal-pqc@nist.gov>; Daniel Smith (b) (6) ; Dang, Thinh H. (Fed) <thinh.dang@nist.gov>
> **Subject:** Re: Kyber's response discussion tomorrow ?
>
> I think we can discuss via email.
>
> I don't think we need to have a meeting tomorrow. Maybe on Tuesday.
>
> Let me know if you think otherwise.

Dustin

From: Dang, Quynh H. (Fed) <quynh.dang@nist.gov>
Sent: Thursday, June 4, 2020 1:34 PM
To: Moody, Dustin (Fed) <dustin.moody@nist.gov>
Cc: internal-pqc <internal-pqc@nist.gov>; Daniel Smith (b) (6) ; Dang, Thinh H. (Fed) <thinh.dang@nist.gov>
Subject: Kyber's response discussion tomorrow ?

Hi Dustin,

Are we going to discuss Kyber's response tomorrow at 10 ?

Quynh.