From: Jacob Alperin-Sheriff <jacobmas@gmail.com>
Sent: Friday, January 05, 2018 12:56 PM
To: pqc-comments
Cc: pqc-forum@list.nist.gov
Subject: OFFICIAL COMMENT: RLCE-KEM

The author states in the abstract of their specification document that "
This document specifies the key encapsulation mechanism RLCE-KEM, where RLCE is a random linear code based McEliece encryption scheme."

and in Section 1, page 4 states that
"It is believed that RLCE security depends on the NP-hardness of decoding random linear codes."

Ignoring quibbles such as the common misnomer than NP-hardness (i.e. worst-case hardness) is directly relevant to cryptography, which requires average-case hardness and vagueness (under what error distribution?), I don't see anything in the specification document that attempts to justify a connection to the hardness of decoding random linear codes.

Such a justification is absolutely needed, as (unlike, say, the trapdoor matrices in public keys in lattice-based cryptography following the Miccancio-Peikert 2012 constructions) the public key G for the RLCE scheme can easily be seen to be statistically distinguishable from random via a simple counting argument.

In particular, there are q^{k^2} choices for S, q^n*(q binom n) choices for G_s, n! for P_1, q^{kw} for the w random r_i, (q-1)^{4w} choices for A and (n+w)! choices for P_2.

Multiplying these all together and using some simple upper bounds, I get the following number of bits of entropy for each of the 6 (real) parameter sets, which I am comparing (on the right) with the number of bits of entropy in a truly random G.

0:	777477	3713000
1:	380383	2361280
2:	1843406	9443040
3:	921036	6118200
4:	5000745	16896000
5:	2447274	11326700

Obviously this doesn't give an efficient distinguisher, but it does show that it's not obvious that G is computationally indistinguishable from random (since it is statistically distinguishable).

In particular, it appears that the only change from a previous insecure scheme is an additional A matrix in the public key, which indeed appears to defeat a direct application of the previous attacks on GRS-code based cryptography given by Couvreur et al in "Distinguisher-based attacks
on public-key cryptosystems using Reed–Solomon codes."

This doesn't seem to be remotely enough to try to claim a connection to the hardness of decoding random linear codes. If I am missing something I would appreciate clarification.

--
-Jacob Alperin-Sheriff

From: Yongge Wang <yongge.wang@gmail.com>
Sent: Friday, January 05, 2018 4:47 PM
To: Jacob Alperin-Sheriff
Cc: pqc-comments; pqc-forum
Subject: Re: [pqc-forum] OFFICIAL COMMENT: RLCE-KEM

Dear Jacob,
thanks for your comments on RLCE-KEM. Regarding your first comment on average hardness security. I believe your comment applies to
almost all code-based schemes submitted to NIST (except for those that is based on Alekhnovich's style construction---but we have
to be very careful on Alekhnovich's style construction based schemes, at least we have one example that needs to be fixed in NIST submissions).

As you mentioned, our sentence is "It is believed that RLCE security depends on the NP-hardness of decoding random linear codes.". So
our analysis is based on heuristic analysis ("belief") and we do not have a strict mathematical proof (again, this is the case for almost all code
based submissions to NIST which do not have security reductions to NP-hard probems).

The submitted document contains more analysis. Let me summarize our justification in a high level here:

1. if we choose n=w, then essentially all columns are randomized.
2. In order to reduce the key size, our parameters for scheme 0,2,4 uses w=n-k. That is, we randomize n-k columns. Due to the fact that we can choose a matrix S to get any k-columns of the generator matrix to be a k \times k identity matrix. We essentially randomize all columns there.
3. In order to further reduce the key size, we use smaller w<n-k for schemes 1,3,5. These selections are specific to the GRS code that we use. These parameters will defeat filtration attacks. If the underlying code is not GRS code, then we may further reduce the value of w so the key size could be further reduced. Due to the limited time, we do not carry out these analysis.

Your analysis on the entropy to distinguish the public key from a truly random G is nice... it shows at least from this viewpoint, one cannot build a distinguisher for RLCE-KEM public key. But we may also note that a distinguisher may not be sufficient to break
a scheme. For example, for some Goppa code based McEliece scheme, we do have distinguishers, but the community still believes
the scheme is secure..

We also note that NIST mentioned that theoretical proof for the security is preferred though not mandatory.

for your last comments:

>This doesn't seem to be remotely enough to try to claim a connection to the hardness of decoding random linear codes. If I am missing something I would appreciate clarification.

We do not claim a proof there. Our sentence "we believe it is related to NP-hardness of decoding random codes" is based on our heuristic analysis.
I hope this will clarify this situation here. Thanks again!

Thanks!
Yongge

On Fri, Jan 5, 2018 at 8:56 PM, Jacob Alperin-Sheriff <jacobmas@gmail.com> wrote:

The author states in the abstract of their specification document that "
This document specifies the key encapsulation mechanism RLCE-KEM, where RLCE is a random
linear code based McEliece encryption scheme."

and in Section 1, page 4 states that
"It is believed that RLCE security depends on the NP-hardness
of decoding random linear codes."

Ignoring quibbles such as the common misnomer than NP-hardness (i.e. worst-case hardness) is directly relevant to cryptography, which requires average-case hardness and vagueness (under what error distribution?), I don't see anything in the specification document that attempts to justify a connection to the hardness of decoding random linear codes.

Such a justification is absolutely needed, as (unlike, say, the trapdoor matrices in public keys in lattice-based cryptography following the Miccancio-Peikert 2012 constructions) the public key G for the RLCE scheme can easily be seen to be statistically distinguishable from random via a simple counting argument.

In particular, there are q^{k^2} choices for S, q^n*(q binom n) choices for G_s, n! for P_1, q^{kw} for the w random r_i, (q-1)^{4w} choices for A and (n+w)! choices for P_2.

Multiplying these all together and using some simple upper bounds, I get the following number of bits of entropy for each of the 6 (real) parameter sets, which I am comparing (on the right) with the number of bits of entropy in a truly random G.

0:	777477	3713000
1:	380383	2361280
2:	1843406	9443040
3:	921036	6118200
4:	5000745	16896000
5:	2447274	11326700

Obviously this doesn't give an efficient distinguisher, but it does show that it's not obvious that G is computationally indistinguishable from random (since it is statistically distinguishable).

In particular, it appears that the only change from a previous insecure scheme is an additional A matrix in the public key, which indeed appears to defeat a direct application of the previous attacks on GRS-code based cryptography given by Couvreur et al in "Distinguisher-based attacks
on public-key cryptosystems using Reed–Solomon codes."

This doesn't seem to be remotely enough to try to claim a connection to the hardness of decoding random linear codes. If I am missing something I would appreciate clarification.

From: Alex Elsayed <eternaleye@gmail.com>
Sent: Friday, January 05, 2018 4:53 PM
To: Yongge Wang
Cc: Jacob Alperin-Sheriff; pqc-comments; pqc-forum
Subject: Re: [pqc-forum] OFFICIAL COMMENT: RLCE-KEM

On Jan 5, 2018 1:47 PM, "Yongge Wang" <yongge.wang@gmail.com> wrote:

> Dear Jacob,
> thanks for your comments on RLCE-KEM. Regarding your first comment on average hardness security. I believe your comment applies to
> almost all code-based schemes submitted to NIST (except for those that is based on Alekhnovich's style construction---but we have
> to be very careful on Alekhnovich's style construction based schemes, at least we have one example that needs to be fixed in NIST submissions).
>
> As you mentioned, our sentence is "It is believed that RLCE security depends on the NP-hardness of decoding random linear codes.". So
> our analysis is based on heuristic analysis ("belief") and we do not have a strict mathematical proof (again, this is the case for almost all code
> based submissions to NIST which do not have security reductions to NP-hard probems).

The problem here isn't your phrasing, it's that "NP-hard" is fundamentally a worst-case notion, not an average-case one.

It's valid to say "hardness" here (as a shorthand for average-case hardness), but not "NP-hardness" (which is a worst-case notion).

From: Yongge Wang <yongge.wang@gmail.com>
Sent: Friday, January 05, 2018 4:57 PM
To: Alex Elsayed
Cc: Jacob Alperin-Sheriff; pqc-comments; pqc-forum
Subject: Re: [pqc-forum] OFFICIAL COMMENT: RLCE-KEM

Alex,

> The problem here isn't your phrasing, it's that "NP-hard" is fundamentally a worst-case notion, not an average-case one.
> It's valid to say "hardness" here (as a shorthand for average-case hardness), but not "NP-hardness" (which is a worst-case notion).

OK.. i see your points there.. I should have pay more attention to these inaccurate terms:-)
Yongge

On Sat, Jan 6, 2018 at 12:53 AM, Alex Elsayed <eternaleye@gmail.com> wrote:

On Jan 5, 2018 1:47 PM, "Yongge Wang" <yongge.wang@gmail.com> wrote:

> Dear Jacob,
> thanks for your comments on RLCE-KEM. Regarding your first comment on average hardness security. I believe your comment applies to
> almost all code-based schemes submitted to NIST (except for those that is based on Alekhnovich's style construction---but we have
> to be very careful on Alekhnovich's style construction based schemes, at least we have one example that needs to be fixed in NIST submissions).
>
> As you mentioned, our sentence is "It is believed that RLCE security depends on the NP-hardness of decoding random linear codes.". So
> our analysis is based on heuristic analysis ("belief") and we do not have a strict mathematical proof (again, this is the case for almost all code
> based submissions to NIST which do not have security reductions to NP-hard probems).

The problem here isn't your phrasing, it's that "NP-hard" is fundamentally a worst-case notion, not an average-case one.

It's valid to say "hardness" here (as a shorthand for average-case hardness), but not "NP-hardness" (which is a worst-case notion).

From: Yongge Wang <yongge.wang@gmail.com>
Sent: Tuesday, April 10, 2018 10:43 PM
To: pqc-comments
Cc: pqc-forum
Subject: Re: OFFICIAL COMMENT: RLCE-KEM

> The attack by Couvreur-Lequesne-Tillich is for the group 1 schemes.
> The attack works due to the fact that I used an inaccurate dimension caculation
> for the filtration attack analysis of group 1 schemes. That is, the formula (31) in the
> RLCEspec.pdf is inaccurate. Their attack will work when w is not close to n-k.

Sorry, here i should say: their attack will work when w<n-k-C where C is a (kind of) constant.
the details could be found in the full descriptions that the team will post later.
thanks!
Yongge

On Tue, Apr 10, 2018 at 10:27 PM, Yongge Wang <yongge.wang@gmail.com> wrote:

For those who attended today's PQC Rump Session, we heard the nice presentation by the team: "Alain Couvreur, Matthieu Lequesne, and Jean-Pierre Tillich" regarding RLCE-KEM (short key version of my submission). Special thanks to the team for communicating an early draft of the attack to me last week. Here I do confirm that their attack works on one group of parameters of the RLCE-KEM submission but will not work on the other group of conservative parameters.

In the RLCE-KEM submission, we have two group's of parameters:

Group1: w<n-k (the scheme with ID=1, 3, 5 and the implementation RLCE_KEM_128A, RLCE_KEM_192A, RLCE_KEM_256A)

Group 2: w=n-k (the scheme with ID=0, 2, 4 and the implementation RLCE_KEM_128B, RLCE_KEM_192B, RLCE_KEM_256B)

The attack by Couvreur-Lequesne-Tillich is for the group 1 schemes.
The attack works due to the fact that I used an inaccurate dimension caculation for the filtration attack analysis of group 1 schemes. That is, the formula (31) in the RLCEspec.pdf is inaccurate. Their attack will work when w is not close to n-k.

This attack will NOT work when w=n-k. That is, the attack will not work fagainst the schemes ID=0, 2, 4 (corresponding to the implementation RLCE_KEM_128B, RLCE_KEM_192B, and RLCE_KEM_256B).

I think the team Couvreur-Lequesne-Tillich will post their attacks here some time in the near future. Thanks Couvreur-Lequesne-Tillich for their analysis of RLCE-KEM. Their analysis also confirms the security of the group 2 parameters (in certain sense).

Thanks!
Yongge