A Comment on Leurent and Thomsen work - New Distinguisher on BMW compression function

The Blue Midnight Wish team

November 16, 2010

Abstract

We give a comment on latest work by Leurent and Thomsen: "New Distinguisher on BMW compression function" [1]. We think that Laurent-Thomsen work is a great result in the study of the compression function of Blue Midnight Wish. However, we also think that the correct title of their work that is compliant with the widely accepted cryptographic terminology should be "Practical Partial-Pseudo-Collisions on the Compression Function of Blue Midnight Wish". This attack fits perfectly in the established framework for analyzing Blue Midnight Wish which we have posted on the SHA-3 forum list on 27/08/2010. Further on, we disagree with Laurent and Thomsen allegation that their work in finding partial pseudo-collisions in Blue Midnight Wish is analogous with the work of den Boer and Bosselaers on MD5 in [2] because den Boer and Bosselaers found **complete** pseudo-collisions on the compression function of MD5, while Leurent and Thomsen found a **partial** pseudo-collision in the compression function of Blue Midnight Wish with 212 bits in the output left out of reach of their controlling technique and because of two essentially different design principles in Blue Midnight Wish that are not present in MD5: Blue Midnight Wish is a double-pipe hash design and is similar with the highly respected cryptographic primitive HMAC. These two design principles renders out all pseudo-attacks (as the one of Leurent and Thomsen) on Blue Midnight Wish as attacks without a potential and a perspective neither to harm nor to break the algorithm.

1 Credits to the work of Leurent and Thomsen

We commend Leurent and Thomsen for their research efforts in connection with their cryptanalysis of the Blue Midnight Wish algorithm. We think that the Leurent-Thomsen paper [1] presents a significant result in the study of the compression function of the Blue Midnight Wish.

We appreciate this new advantage in the non-trivial study of differential properties of the compression function of Blue Midnight Wish because this is important and very difficult to achieve. Despite of incorrect conclusions in the paper, it is a great contribution to the study of the differential properties of Blue Midnight Wish compression function. The paper shows new and very nice way how to manipulate differentials inside the two thirds of the compression function. It also shows how difficult it is to bypass entangling bijections, used in the compression function, what is one of the basic building principle of Blue Midnight Wish. We thank Leurent and Thomsen for this great work.

2 Critique of some of the claims and alleged implications of their work to the security of the Blue Midnight Wish hash function

We organize our critical remarks in 5 points.

1. The latest attack on the compression function of Blue Midnight Wish hash function by Leurent and Thomsen is again a pseudo-attack since they control both the message and the chaining value. Thus, the correct title that is compliant with the widely accepted cryptographic terminology should be "*Practical Partial-Pseudo-Collisions on the Compression Function of* Blue Midnight Wish".

2. This attack fits perfectly in the established framework for analyzing Blue Midnight Wish which we have posted on the SHA-3 forum list on 27/08/2010 [3]. The partial pseudo-collision that they find has three fully collided values in the first part of the chaining value. Specifically, their pseudo-attack achieves complete collision on the first 3 variables of the chaining value and partial collision on 7 additional variables, leaving 6 variables of the output beyond collision control. The updated framework which includes the latest pseudo-attack of Leurent and Thomsen is already included in our web page: `http://www.q2s.ntnu.no/sha3_nist_competition/start` and the corresponding pdf: `http://people.item.ntnu.no/~danilog/Hash/BMW-SecondRound/FrameworkHowToEvaluateSecurityInBMW-Nov-2010.pdf` may be downloaded from there.

3. Leurent and Thomsen seem to be trying to increase the value of their analysis by giving their work a perspective and potential impact similar to that of den Boer and Bosselaers' work on their MD5 analysis. From the Leurent and Thomsen paper, we quote: "*To put such attacks into perspective, one might look at the attacks on MD5. The first attack on the compression function was found in 1993 by den Boer and Bosselaers [5], using a very simple differential path. This attack did not threaten the iterated hash function, but the path used in the attack is a core element of the successful attack of Wang et al. in 2005 [10].*"

   There are at least two evident mismatches in using the analogy between the work of den Boer and Bosselaers [2] and the history of the MD5 analysis and the work of Leurent and Thomsen on Blue Midnight Wish:

   a) The collisions for the compression function of MD5 which were found by den Boer and Bosselaers were described by a precise terminology as pseudo-collisions by Robshaw in [4] and by Dobbertin in [5]. Leurent and Thomsen should also strive to use such precise terminology

   b) den Boer and Bosselaers found **COMPLETE** pseudo-collisions on the narrow-pipe compression function of MD5, while Leurent and Thomsen found a **PARTIAL** pseudo-collision in the double-pipe compression function of Blue Midnight Wish, with 212 bits in the output left out of reach of their controlling technique.

   However, since they have put their work into this perspective and are making allegations that their work will decrease the confidence in Blue Midnight Wish as den Boer's and

Bosselaer's work did for MD5, we would also like to put into perspective their attack (and all other pseudo-attacks) on BLUE MIDNIGHT WISH by recalling the similarity of the finalization of the BLUE MIDNIGHT WISH algorithm with the HMAC (that fact Leurent and Thomsen are mentioning in the introduction to their work and was first mentioned in the analysis of the SHA-3 candidates done by Andreeva, Mennink and Preneel in [6]). So having a hash function for which similar design principles as for HMAC (one of the most trusted designs in the contemporary cryptology) have been used, clearly increases the confidence in BLUE MIDNIGHT WISH and renders out all pseudo-attacks on it as attacks without a potential and a perspective neither to harm nor to break the algorithm.

4. We do not see as a truthful and as a big achievement their claim in the conclusion "*We also note that if the compression function is truncated like in the final transformation of BMW, we can still build pairs of message which collide in more than 110 bits with complexity* 2^{32}*. This is the first distinguisher on the truncated compression function of BMW.*" As already noted, it is not a distinguisher but a pseudo-distinguisher. With the same computational effort that they are using (2^{32} calls to the compression function) a generic partial collision search can find a real partial collision on approximately 198 bits on the second-half of the chaining value (by "real" we mean without the need to control every input into the compression function i.e. without going into the direction of a pseudo-attack).

   Moreover, the claim: "*... if the compression function is truncated like in the final transformation of* BLUE MIDNIGHT WISH*, we can still build pairs of messages which collide in more than 110 bits with complexity* 2^{32}*.*" is not correct. In the part where they say that they are able to build "*pairs of messages*", the precise phrasing would be ..."*pairs of new chaining values and pairs of final constants M ...*", because M is no longer a message block in the final transformation of BLUE MIDNIGHT WISH.

   Additionally, the presented pseudo-distinguisher requires a huge control of the chaining variable H which in the final transformation is a pre-computed value obtained by digesting the message and in Laurent-Thomsen work there are no indications how their complete control over the chaining variable and the message blocks can be transformed into an attack that controls the whole message. If the conclusion is rewritten correctly, we would see the following statement: "*We also note that if the output of the compression function is truncated to its half, we can still build pairs of constants and pairs of chaining-hash values which collide in more than 110 bits with complexity* 2^{32}*.*" Now this is true, but without any value. As we have already stated, a much better partial-collision result is possible to obtain without employing the pseudo-attack by a simple generic search of partial collisions. So the conclusion in their paper should be corrected in some way (and if we were given their draft work in advance as it is a general ethical attitude in academic Cryptologic research) we would have pointed out these incorrect parts.

5. Let us analyze the implication of finding partial pseudo-collisions on the security of the hash function. Recall that the final step of the hash function BLUE MIDNIGHT WISH is $C(H_{LAST}, CONST)$, where H_{LAST} is the value of the previous iterative hashing of the padded message m and $C()$ is the compression function. As it was correctly noted

in the paper, it is similar to HMAC construction - H_{LAST} is not a message block now, but a "*pre-hash*" value. Moreover, in the case of Blue Midnight Wish, the length of this pre-hash value is twice as long as in the case of HMAC construction!

Let us suppose the attacker succeeded to find a collision or a near-collision on the whole Blue Midnight Wish hash function. How he/she succeeded to do that? There are only two cases. The first one is that the values $H_{LAST} = H'_{LAST}$ in the last step are the same and are coming from two different digested messages $m \neq m'$. The second one is that the values H_{LAST} and H'_{LAST} in the last step are different.

a) In the first case the attacker found a complete collision (not a pseudo-collision) of the compression function (with double length output). So, in this case, the first necessary condition is that the attacker has to find a **COMPLETE** collision of the double-pipe compression function. Note that it is only necessary, not sufficient condition, because there has to be a way how to obtain this value $H_{LAST} = H'_{LAST}$ for two different messages from the beginning of hashing. The important note is that any near-collision even on all bits but one is not useful. It has to be complete collision on all bits of the double-pipe compression function! Usually, finding near-collision of the compression function is a great result. Here it could be even contra productive. Having very near values H_{LAST} and H'_{LAST} i.e. $Hamming(H_{LAST}, H'_{LAST})$ is low, the final operation $C(H_{LAST}, CONST)$ and $C(H'_{LAST}, CONST)$ will diffuse them into two values having Hamming distance around 256 (512 for BMW512). Bijections used in Blue Midnight Wish behave like MDS codes - small changes in the input guarantee big changes in the output. So the first task is different from the traditional hash constructions: we need **COMPLETE** collision of the compression function, *assuming that ANY NEAR-COLLISION IS VERY NEGATIVE RESULT*! Moreover the first task *works on DOUBLE LENGTH* values compared with the traditional narrow-pipe hash designs.

How to measure the effectiveness of Leurent-Thomsen near-collision of the compression function? Should we continue to extend their near-collisions from 300 to more bits up to 511? We have just offered arguments that from the breaking point of view for the whole hash function it could be even contra productive. But it has a big sense and big importance for understanding the insights of the hash function and to study its properties. So their paper is great, because it shows some properties of the compression function. This paper shows that even with a total control of every input variable in the compression function of Blue Midnight Wish, at best you can get is a partial collision which has not much use for breaking the real hash function, and that speaks very much in favor of the strength of the hash function. Just compare the situation of having a total control over all inputs of the "compression function" of the sponge designs - you need only 2 calls to the inverse of the bijective function and you have a **COMPLETE pseudo-collision** (not that it has anything to do with the general strength of the sponge-based hash designs).

b) The alternative to the first case is the second case consisting of finding two different pre-hash values $H_{LAST} \neq H'_{LAST}$ such that the $chopC(H_{LAST}, CONST)$ and

$chopC(H'_{LAST}, CONST)$ are equal or near. Note again that this is not sufficient condition for a successful attack, because the attacker in this case have to find a way how to obtain these pre-hash values $H_{LAST} \neq H'_{LAST}$ for two different messages from the beginning of hashing. In this second task the attacker has to explore the function $H_{LAST} \mapsto chopC(H_{LAST}, CONST)$. This function is very different from the compression function $(M, H) \mapsto C(M, H)$ since the roles of M and H are now swapped. Moreover, the partial transformations inside the *chopC* are very different from $C(M, H)$, when one variable is a constant.

And, of course, the function $H_{LAST} \mapsto chopC(H_{LAST}, CONST)$ has half freedom both in input and output variables, so differential strategies and paths will be very different from the first case. Also, when you look at the nice Fig.1 of the Leurent-Thomsen paper, the variable M is now going into H_{LAST}, the variable Q_a is now (due to the constant $CONST$) a **BIJECTIVE image** of H_{LAST}, and Q_b is some kind of one-way function of H_{LAST}. These three variables are inputs to the function f_2. Now, Q_a behaves like MDS code of H_{LAST} - the smaller changes in H_{LAST}, the bigger changes in $Q_a(H_{LAST})$, so in the couple $(H_{LAST}, Q_a(H_{LAST}))$ there is guaranteed some amount of changes in total. The behavior of the special function $H_{LAST} \mapsto f_2(H_{LAST}, Q_a(H_{LAST}), Q_b(H_{LAST}))$ is crucial. This is the second way how to explore collisions of Blue Midnight Wish, which has not been explored so far. We would like to stimulate any research in this direction.

We can conclude this point that whatever the attacker knows and uses, he/she has to complete either scenario a) or scenario b). In the first scenario it is necessary to obtain the complete collision on the double pipe. Obtaining near-collision has no significance for launching an attack, it has a meaning for the study of the compression function. And this is the case of Leurent-Thomsen paper.

References

[1] G. Leurent and S. S. Thomsen, "Practical Partial-Collisions on the Compression Function of BMW", SHA-3 Hash forum list from 12 Nov 2010.

[2] B. den Boer and A. Bosselaers, "Collisions for the compression function of MD5", Advances in Cryptology Eurocrypt 93, LNCS, vol. 773, Springer-Verlag, 1994, pp. 293-304.

[3] Blue Midnight Wish team, "A framework for Measuring and Evaluating the Progress of the Cryptanalysis of the Hash Function Blue Midnight Wish", SHA-3 Hash forum list from 27 Aug 2010.

[4] M. Robshaw, "On pseudo-collisions in MD5", Technical Report TR-102, ver. 1.1., RSA Laboratories, July 1994.

[5] H. Dobbertin, "Cryptanalysis of MD5 compress", presented at the rump session of EurocrCrypt'96.

[6] E. Andreeva and B. Mennink and B. Preneel, "Security Reductions of the Second Round SHA-3 Candidates", Cryptology ePrint Archive, Report 2010/381.