On the Randomness and Regularity of Reduced EDON-$\mathcal{R}$ Compression Function

Rune Steinsmo Ødegård[1] **and Danilo Gligoroski**[2]
[1]Centre for Quantifiable Quality of Service in Communication Systems, NTNU, Trondheim, Norway
[2]Department of Telematics, NTNU, Trondheim, Norway

May 25, 2009

Abstract

EDON-$\mathcal{R}$ is one of the candidate hash functions for the ongoing NIST competition for the next cryptographic hash standard called SHA-3. Its construction is based on algebraic properties of non-commutative and non-associative quasigroups of orders 2^{256} and 2^{512}. In this paper we are giving some of our results in investigation of the randomness and regularity of reduced EDON-$\mathcal{R}$ compression functions over quasigroups of order 2^8 and 2^{16}. Our experiments show that the Bellare-Khono balance of EDON-$\mathcal{R}$ compression function is high. Actually, for the reduced EDON-$\mathcal{R}$ with quasigroups of order 2^8 we show that the compression function is perfectly balanced, while with quasigroups of order 2^{16} the Belare-Khono balance is $\mu(R_{16}) = 0.99985$.

Keywords: hash function, randomness, regularity, balance

1 Introduction

Recently Gligoroski et.al submitted the hash function EDON-$\mathcal{R}$ [5] to the NIST hash competition [2]. With speeds of 5.77 cycles/byte and 3.15 cycles/byte on amd64 1401MHz Intel Core 2 Duo for EDON-$\mathcal{R}$256 and EDON-$\mathcal{R}$512 respectively, EDON-$\mathcal{R}$ is the fastest hash function submitted to the competition [1]. This has generated a lot of attention in the cryptographic community and much effort has therefore been put into breaking EDON-$\mathcal{R}$. So far there have been various limited results. Klima [7] showed the possibility of 2^K multicollisions requiring $K * 2^{n/2}$ computations and access to $2^{n/2}$ units of memory. Khovratovich et.al [6] noted the possibility of free-start collisions and used this to launch a preimage attack on EDON-$\mathcal{R}$ requiring $2^{2n/3}$ computations and access to $2^{2n/3}$ units of memory. Later Gligoroski and Ødegård [4] disputed the validity of the model in which the attack of Khovratovich et. al is compared to generic attacks. The latest result on EDON-$\mathcal{R}$ by Leurent [8] showed the possibility of key recovery using $2^{5n/8}$ operations when EDON-$\mathcal{R}$ is used as a special secret prefix MAC. Note that all identified weaknesses of EDON-$\mathcal{R}$ are only present in the free start collision case. In general, the problem of free start collisions can be addressed for instance by the Davies-Meier method for feed-forwarding of the chaining value.

The design of EDON-$\mathcal{R}$ is a double piped iterated compression function. As part of our cryptanalysis of the EDON-$\mathcal{R}$ hash function we present here results from tests of randomness performed on EDON-$\mathcal{R}$ compression functions reduced in size.

Using the same theory as [5], we constructed an 8-bit EDON-$\mathcal{R}$ compression function. This construction is small enough that we can test all 2^{32} possible inputs of messages and chaining values. The distribution of the output was then compared to what is expected from an ideal random function. In addition we tested to see if the function is regular.

Similarly, we also constructed a 16-bit EDON-$\mathcal{R}$ compression function. For 300 different chaining values chosen at random, we tested all 2^{32} possible message inputs. The distribution of the output was then compared to what is expected from an ideal random function. We also used the results to compute the balance of the 16-bit compression function.

This paper is organized as follows. We first give the required Background in Section 2. In Section 3 and Section 4 we construct and analyze 8-bit and 16-bit EDON-$\mathcal{R}$ respectively. Finally Section 5 concludes the paper.

2 Background

In this section we give the required mathematical background for this paper. The underlying structure of EDON-$\mathcal{R}$ are quasigroups of order 2^{kw} where $kw = 256$ and 512 for EDON-$\mathcal{R}$-256 and EDON-$\mathcal{R}$-512 respectively.

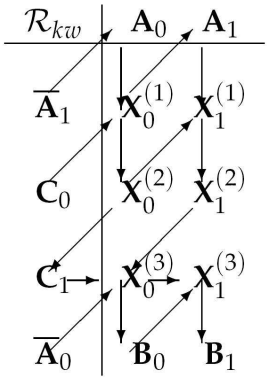

Figure 1: Schematic representation of the function $\mathcal{R}_{kw}$. The diagonal arrows can be interpreted as quasigroup operations between the source and the destination, and the vertical or the horizontal arrows as equality signs "=".

Definition 1 A *quasigroup* $(Q, *)$ is an algebraic structure consisting of a nonempty set Q and a binary operation $* : Q^2 \to Q$ with the property that each of the equations

$$\begin{aligned} a * x &= b \\ y * a &= b \end{aligned} \tag{1}$$

has unique solutions x and y in Q.

The compression function $\mathcal{R}_{kw}$ of EDON-$\mathcal{R}$ is a series of quasigroup operations of the form

$$\mathbf{X} *_{kw} \mathbf{Y} = \pi_{1,kw}(\pi_{2,kw}(\mathbf{X}) + \pi_{3,kw}(\mathbf{Y})), \tag{2}$$

where the permutations $\pi_{i,kw}$ treat $\mathbf{X}, \mathbf{Y}$ as vectors of $k = 8$ words of size $w = 32, 64$ bits. The permutation $\pi_{1,kw}$ is a simple reordering of the variables. The permutations $\pi_{2,kw}$ and $\pi_{3,kw}$ are defined from two orthogonal Latin squares of size $k = 8$ (the same size as the vectors). For a detailed explanation of how these permutations are defined from the Latin squares we refer the reader to [5].

Definition 2 A *Latin square* of size k is an $k \times k$-matrix whose elements are the numbers $0, \ldots, k-1$ and each number appears exactly one time in each row and each column.

The compression function is defined by repeated use of the quasigroup operation as shown in Figure 1. This gives the following formula for the chaining values $(\mathbf{B}_0, \mathbf{B}_1)$

$$\mathcal{R}_{kw}(\mathbf{C}_0, \mathbf{C}_1, \mathbf{A}_0, \mathbf{A}_1) = (\mathbf{B}_0, \mathbf{B}_1) \tag{3}$$

where

$$\begin{aligned} \mathbf{B}_0 &= \overline{\mathbf{A}}_0 * ((\mathbf{C}_1 * (\overline{\mathbf{A}}_1 * \mathbf{A}_0)) * \mathbf{C}_0) \\ \mathbf{B}_1 &= (\overline{\mathbf{A}}_0 * ((\mathbf{C}_1 * (\overline{\mathbf{A}}_1 * \mathbf{A}_0)) * \mathbf{C}_0)) * \\ &\quad (((\mathbf{C}_1 * (\overline{\mathbf{A}}_1 * \mathbf{A}_0)) * ((\overline{\mathbf{A}}_1 * \mathbf{A}_0) * \mathbf{A}_1)) * \\ &\quad ((\mathbf{C}_1 * (\overline{\mathbf{A}}_1 * \mathbf{A}_0)) * \mathbf{C}_0)). \end{aligned} \tag{4}$$

Definition 3 A function is said to be *regular* if every possible output has 2^m preimages for some m.

In [3] Bellare and Kohno looked at the connection between the "amount of regularity" of a hash function and the general $2^{n/2}$ bound for birthday attacks on hash functions of size n. Their conclusion was that $2^{n/2}$ was the lower bound only if the hash function was regular, and that the actual lower bound can be significantly less depending on the regularity of the hash function. Bellare and Khono introduced *balance* as a measure of regularity and used its properties to prove the relation between balance and the expected number of trials in the birthday attack.

$$L_1 = \left[\begin{array}{cccc} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ \hline 3 & 2 & 1 & 0 \end{array}\right] \qquad L_2 = \left[\begin{array}{cccc} 0 & 1 & 2 & 3 \\ 2 & 3 & 0 & 1 \\ 3 & 2 & 1 & 0 \\ \hline 1 & 0 & 3 & 2 \end{array}\right]$$

Table 1: Two mutually orthogonal Latin squares used to define the permutations $\pi_{2,8}$ and $\pi_{3,8}$

Quasigroup operation of order 2^8

Input: $\mathbf{X} = (X_0, X_1, X_2, X_3)$ and $\mathbf{Y} = (Y_0, Y_1, Y_2, Y_3)$
where X_i and Y_i are 2–bit variables.
Output: $\mathbf{Z} = (Z_0, Z_1, Z_2, Z_3)$ where Z_i are 2–bit variables.
Temporary 2–bit variables: $T_0, \ldots, T_7$.

1.
$$\begin{array}{llll} T_4 & \leftarrow & ROTL^0(\texttt{0x1} & + \ X_0 + X_1 + X_2; \\ T_5 & \leftarrow & & ROTL^1(\ X_0 + X_1 + X_3; \\ T_6 & \leftarrow & & ROTL^0(\ X_0 + X_2 + X_3; \\ T_7 & \leftarrow & & ROTL^1(\ X_1 + X_2 + X_3; \end{array}$$

2.
$$\begin{array}{llll} T_0 & \leftarrow & ROTL^0(\texttt{0x2} & + \ Y_0 + Y_2 + Y_3; \\ T_1 & \leftarrow & & ROTL^1(\ Y_1 + Y_2 + Y_3; \\ T_2 & \leftarrow & & ROTL^0(\ Y_0 + Y_1 + Y_2; \\ T_3 & \leftarrow & & ROTL^1(\ Y_0 + Y_1 + Y_3; \end{array}$$

3.
$$\begin{array}{llll} Z_3 & \leftarrow & T_7 + T_1; \\ Z_2 & \leftarrow & T_6 + T_0; \\ Z_0 & \leftarrow & T_5 + T_2; \\ Z_1 & \leftarrow & T_4 + T_3; \end{array}$$

Table 2: An algorithmic description of the quasigroup operation of order 2^8.

Definition 4 Let $h : R \rightarrow D$ be a function whose domain D and range $R = \{R_1, \ldots, R_r\}$ have sizes $d, r \geq 2$, respectively. For $i \in \{1, \ldots, r\}$ let $d_i = |h^{-1}(R_i)|$ denote the size of the pre-image of R_i under h. The balance of h, denoted $\mu(h)$, is defined as

$$\mu(h) = \log_r \left[\frac{d^2}{d_1^2 + \cdots + d_r^2}\right] \tag{5}$$

where $\log_r(\cdot)$ denotes the logarithm in base r.

Note that Bellare and Khono showed that a Merkle-Damgård transform does not necessarily preserve balance. This means that in addition to the balance of the compression function, the balance of the whole hash function should also be investigated [3].

3 Analysis of 8-bit EDON-$\mathcal{R}$

To test some of the properties of the function $\mathcal{R}_{kw}$ we constructed a small version using the same theory. Setting the size of the vectors to $w = 4$ and the size of the words to $k = 2$, an 8-bit version was constructed using the two orthogonal Latin squares in Table 1. Alternating between left rotation of 0 and 1 we arrived at the quasigroup operation in Table 2

One notable difference between this quasigroup operation and the ones defined in [5] is the missing XOR parts. The reason for this difference is that here there is only one row below the line in the Latin squares used to define the permutations. So instead of XORing the variables, they are permuted according to the rows $(3, 2, 1, 0)$ and $(1, 0, 2, 3)$.

3.1 Experiments and results for 8-bit EDON-$\mathcal{R}$

We have now constructed a reduced EDON-$\mathcal{R}$ compression function $R_8 : 2^{32} \rightarrow 2^{16}$ which is small enough to exhaustively go trough all possible input values. To test some of the properties of this function we performed the following two experiments.

l	Min	Max	Mean	IRF	Difference %
0	23285	24108	23694,53	24109.16	1.72%
1	24007	25102	24538.69	24109.53	-1.78 %
2	11902	12624	12251.78	12054.77	-1.63 %
3	3704	4142	3936.85	4018.19	2.02 %
4	792	1046	919.00	1004.52	8.51 %
5	117	224	167.37	200.89	16.69 %
6	9	47	24,46	33.48	26.94 %
7	0	12	2.99	4.78	37.51 %
8	0	4	0.30	0.60	50.17 %
9	0	3	0.027	0.066	59.56 %
10	0	1	0.0015	0.0066	77.94 %
≥ 11	0	0	0	0.00066	100.00 %

Table 3: The distribution of the image of $\mathcal{R}_8$ for all possible pairs of chaining values. The second last columns show what is expected for an ideal random function (IRF), while the first 3 columns show the result for $\mathcal{R}_8$. The last column show the difference in percent between the IRF and the mean of $\mathcal{R}_8$.

Experiment 1 The first test we performed was on the number of collisions of the compression function. Holding the chaining values $(\mathbf{C}_0, \mathbf{C}_1)$ constant, we used all possible pairs $(\mathbf{M}_0, \mathbf{M}_1)$ of messages as input to the compression function.

$$R_8(\mathbf{C}_0, \mathbf{C}_1, \mathbf{M}_0, \mathbf{M}_1) = (\mathbf{B}_0, \mathbf{B}_1) \tag{6}$$

The output $(\mathbf{B}_0, \mathbf{B}_1)$ represented as a number between 0 and $2^{16} - 1$ was then tallied with respect to collisions. That is, we counted the number of elements mapped to 0 times, the number of elements mapped to 1 times and so on. This was done for all possible pairs of chaining values. The result was then compared to what is expected for an ideal random function (IRF).

If $\mathcal{R}_8$ is an IRF each element in the image should be mapped to by the probability $p = 2^{-16}$. This means that the an element will be mapped to l times out of n with probability

$$P(l) = \binom{n}{l} p^k (1-p)^{n-l}, \tag{7}$$

where $n = 2^{16}$. This means that expected number of elements mapped to l times is $nP(l)$.

Result The result from this experiment is listed in Table 3 and shown in Figure 2. From the table we see that even the very reduced $\mathcal{R}_8$ function has a distribution similar to an IRF. For $l = 0, 1, 2, 3$ the results are good with at most 2% difference from an IRF. However the difference rapidly grows for bigger l. For all l the 95% confidence interval for the mean was outside what is expected for an IRF. This means that the difference between $\mathcal{R}_8$ and an IRF is statistically significant. We also noticed some other non-random behavior. Sorting the result according to the first chaining value $\mathbf{C}_0$ we see that all occurrences of 9 and 10 collisions are centered around certain values of $\mathbf{C}_0$. The same is not true if we sort with respect to $\mathbf{C}_1$. This means the first chaining value has bigger influence on the final output then the second. Looking at the construction of the compression function in Figure 1 this result is of course expected since the first chaining value is introduced earlier in the computation.

Experiment 2 The second test we performed was on the number of pre-images each element has under the compression function. We exhaustively went trough all 2^{32} possible input values and looked at how many times each element in the image was mapped to.

Result The result from this experiment was that each element was mapped to exactly 2^{16} times. This means that the function $\mathcal{R}_8$ is regular according to Definition 3.

4 Analysis of 16-bit EDON-$\mathcal{R}$

We also constructed a 16-bit version of $\mathcal{R}_{kw}$ by setting the size of the vectors to $k = 8$ and the wordsize to $w = 2$. For this construction we used the same Latin squares as in the construction of EDON-$\mathcal{R}$256 and 512. This means that the algorithmic

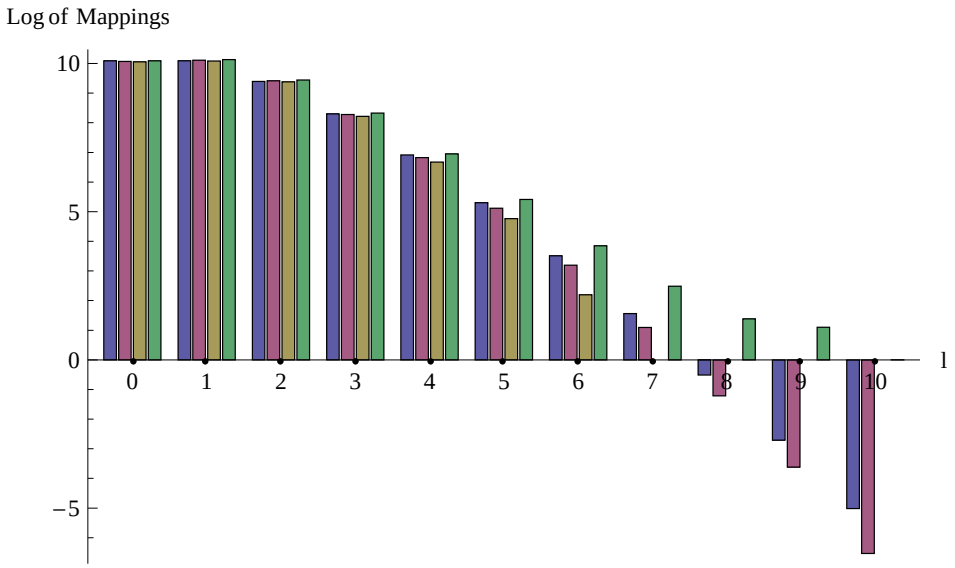

Figure 2: A bar chart comparing the binomial distribution of an IRF with $\mathcal{R}_8$. The x-axis is how many times, l, an element is mapped to, while the y-axis is the logarithm of how many elements is mapped to l times. The first bar is the IRF, the second bar is the mean of $\mathcal{R}_8$, while the third and fourth bar are the minimum and maximum of $\mathcal{R}_8$ respectively.

description of the quasigroup operation of order 2^{16} are very similar to the ones found in [5]. The only difference is that we alternated between rotation of 0 and 1 instead of the rotations described for EDON-$\mathcal{R}$256 and 512.

4.1 Experiment and results for 16-bit EDON-$\mathcal{R}$

The compression function $\mathcal{R}_{16} : 2^{64} \rightarrow 2^{32}$ is to big to go trough all possible input values. However we can still perform an experiment on the distribution of the output similar to Experiment 1 in Section 3.

Experiment 3 Holding the chaining values $(\mathbf{C}_0, \mathbf{C}_1)$ constant we used all possible pairs $(\mathbf{M}_0, \mathbf{M}_1)$ of messages as input to the compression function.

$$R_{16}(\mathbf{C}_0, \mathbf{C}_1, \mathbf{M}_0, \mathbf{M}_1) = (\mathbf{B}_0, \mathbf{B}_1) \tag{8}$$

The output $(\mathbf{B}_0, \mathbf{B}_1)$ represented as a number between 0 and $2^{32} - 1$ was then tallied with respect to collisions. This test was performed for 300 different pairs of chaining values chosen at random.

If $\mathcal{R}_{16}$ is an ideal random function, each element in the image should be mapped to by the probability $p = 2^{-32}$. This means that the an element will be mapped to l times out of n with probability

$$P(l) = \binom{n}{l} p^k (1-p)^{n-l}, \tag{9}$$

where $n = 2^{32}$. Which means that expected number of elements mapped to l times is $nP(l)$.

Result The result from this experiment is listed in Table 4. Note that, because of some inaccuracy[1] in the computation of the distribution of the ideal random function, the sum of the expected results for 0 to 14 is slightly more then 2^{32}. The expected number of collisions larger then 15 is therefor listed as not available. Other approximations show this number to be in the order of 10^{-3}.

From the table we see that the compression function of 16-bit EDON-$\mathcal{R}$ has an output distribution very similar to an IRF. For most values of l the difference between R_{16} and an IRF is much less then 1%. For $l = 11, 12, 13$ and 14 the difference is larger. Some of the reason for this difference is that the probability for 11 or more mappings colliding is very small and the variance for such collisions is therefor higher. This is also reflected in the 95% confidence intervals for the mean. For $l = 1, \ldots, 6$ the computed confidence intervals is outside what is expected for an IRF, while for $l = 7, \ldots, 14$ the confidence intervals for the

[1]The number $(1 - 2^{-32})^{2^{32}-l}$ could only be approximated using Mathematica 6.0.

l	Min	Max	Mean	IRF	Diff %
0	1579952990	1580105601	1580014613	1580030157	0.00098 %
1	1579915274	1580140186	1580047057	1580030350	-0.0011 %
2	789949319	790105082	790021531	790016352	-0.00066 %
3	263295979	263371739	263335106	263337413	0.00088%
4	65813150	65860809	65831618.77	65834545.53	0.0044%
5	13154784	13174678	13165622.68	13166934.63	0.010%
6	2190432	2198973	2194306.50	2194480.82	0.0079%
7	312150	314757	313445.75	313498.08	0.017%
8	38656	39803	39169.01	39187.43	0.047%
9	4139	4563	4349.83	4354.11	0.098%
10	361	491	433.45	435.41	0.45%
11	23	57	39.09	39.58	1.24%
12	0	10	3.46	3.30	-4.99%
13	0	3	0.26	0.25	-3.78%
14	0	1	0.017	0.018	8.041%
≥ 15	0	0	0	NA	NA

Table 4: The distribution of the image of $\mathcal{R}_{16}$ for all possible pairs of chaining values. The second last columns show what is expected for an ideal random function (IRF), while the first 3 columns show the result for $\mathcal{R}_{16}$. The last column show the difference in percent between the IRF and the mean of $\mathcal{R}_{16}$.

mean contains what is expected for an IRF. This means that altough the output distribution of $\mathcal{R}_{16}$ is very similar to an IRF they are still significantly different.

Experiment 4 Because of the size of $\mathcal{R}_{16}$ we were not able to test wether the function is regular. We did however tally how many times of the $300 * 2^{32}$ different input values each element was mapped to. The tally was used to compute the Bellare-Khono balance of $\mathcal{R}_{16}$ as defined in Definition 4.

Result Computing the sum of the square of the number of preimages for each of the 2^{32} possible output values we got the number 387835522366350. This gives the following result for the balance.

$$\mu(\mathcal{R}_{16}) = \log_{2^{32}} \left[\frac{(300 * 2^{32})^2}{387835522366350} \right] = 0.99985 \tag{10}$$

We will also quickly mention some other possibly interesting numbers from this experiment. The least amount of times a number was mapped to was 194, while 413 was the most amount of times a number was mapped to (300 of course being the average). The variance was $\sigma^2 = 299.9943$.

The results for the regularity of $\mathcal{R}_8$ and the balance of $\mathcal{R}_{16}$ together with the analysis for delta deviations in our documentation of EDON-$\mathcal{R}$ [5] are strong evidence that the balance is very high for $\mathcal{R}_{kw}$ in general. An open and interesting question is for what k and w the function $\mathcal{R}_{kw}$ is completely regular.

5 Conclusion

We have shown that the reduced compression function $\mathcal{R}_8$ is regular and that its output distribution is similar to that of an ideal random function. We have also shown that distribution of the output of $\mathcal{R}_{16}$ is very similar to an ideal random function, and that the balance of $\mathcal{R}_{16}$ is high.

Comparing Table 3 and Table 4 we see that the amount of randomness drastically increases as we increase the size of the range and the domain of the compression function. Intuitively we expect this trend to follow as we increase the size of the range and the domain of the compression function even more.

Based on the results of this paper it is difficult to give a general prediction for the correlation between the size of the compression function and its balance. It is possible that also $\mathcal{R}_{16}$ is completely regular, but unfortunately we do not have the computer power to test this. Doing some tests on the balance of $\mathcal{R}_{256}$ and $\mathcal{R}_{512}$ would be quite interesting in this regard.

Additionally, it is clear that our methodology of analyzing EDON-$\mathcal{R}$ by reducing the size of the variables and then investigating the properties of such severely reduced function can be applied to all hash functions.

References

[1] ECRYPT Benchmarking of All Submitted Hashes. measurements of hash functions. `http://bench.cr.yp.to/results-hash.html`.

[2] *Announcing Request for Candidate Algorithm Nominations for a New Cryptographic Hash Algorithm (SHA-3) Family*. NIST, 2007. `http://csrc.nist.gov/groups/ST/hash/index.html`.

[3] Mihir Bellare and Tadayoshi Kohno. Hash function balance and its impact on birthday attacks. In *Advances in Cryptology - EUROCRYPT 04, Lecture Notes in Computer Science*, pages 401–418. Springer-Verlag, 2004.

[4] Danilo Gligoroski and Rune Steinsmo Ødegård. On the complexity of Khovratovich et. al's preimage attack on EDON-$\mathcal{R}$. Available online, 2009.

[5] Danilo Gligoroski, Rune Steinsmo Ødegård, Marija Mihova, Svein Johan Knapskog, Ljupco Kocarev, Aleš Drápal, and Vlastimil Klima. Cryptographic hash function EDON-$\mathcal{R}$. Submission to NIST, 2009.

[6] Dmitry Khovratovich, Ivica Nikolic', and Ralf-Philipp Weinmann. Cryptanalysis of EDON-$\mathcal{R}$. Available online, 2008.

[7] Vlastimil Klima. Multicollisions of EDON-$\mathcal{R}$ hash function and other observations. Available online, 2008.

[8] Gaëtan Leurent. Key recovery attack against secret-prefix EDON-$\mathcal{R}$. Cryptology ePrint Archive, Report 2009/135, 2009. `http://eprint.iacr.org/`.