Analysis of Xorrotation With Application to an HC-128 Variant

Stankovski, Paul; Hell, Martin; Johansson, Thomas

Published in:
LNCS

2012

Link to publication

Citation for published version (APA):
Stankovski, P., Hell, M., & Johansson, T. (2012). Analysis of Xorrotation With Application to an HC-128 Variant. In W. Susilo, Y. Mu, & J. Seberry (Eds.), *LNCS* (Vol. 7372, pp. 419-425). Springer.

Total number of authors:
3

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Analysis of Xorrotation With Application to an HC-128 Variant[*]

Paul Stankovski, Martin Hell, and Thomas Johansson

Dept. of Electrical and Information Technology, Lund University,
P.O. Box 118, 221 00 Lund, Sweden

Abstract. Many cryptographic primitives rely on word rotations (R) and xor (X) to provide proper mixing. We give RX-system mixing a very general treatment and deduce some theoretical results on related probability distributions. Pure RX-systems are easy to break, so we show how to apply our theory to a more complex system that uses RX operations in combination with S-boxes. We construct an impractical (keystream complexity $2^{90.9}$), but new and non-trivial distinguisher for a variant of HC-128 for which modular addition is replaced with xor.

Keywords: RX, probability distribution, stream cipher, HC-128, cryptanalysis, distinguisher

1 Introduction

Consider the Xorrotation family

$$g_{w,r_1,\ldots,r_n}(x) = x \oplus (x \lll r_1) \oplus \ldots \oplus (x \lll r_n)$$

of functions for which x is a w-bit word, $\oplus$ denotes xor and $\lll$ denotes left rotation (cyclic shift) with respect to the word length w. For the rotation amounts r_i we have $0 < r_i < w$. These bit mixing functions are often used in cryptographic primitives to provide intra-word diffusion.

While primitives that rely on modular addition (A), rotation (R) and xor (X) are commonly labeled ARX, $g_{w,r_1,\ldots,r_n}$ is RX. Pure AX- and RX-systems have been shown to be weak, see [3, 8], but we will show how our theory can be used in practice by applying it to a more complex system that includes RX operations *and* S-boxes.

The Xorrotation function family was studied by Thomsen [11] and Rivest [9]. Thomsen showed that the mapping is invertible for all choices of distinct r_i with $0 \leq r_i < w$, and all word lengths $w = 2^k$ if n is odd. Very recently, Rivest gave a different and more general proof, a proof that

[*] The final publication is available at springerlink.com. This is the short version of the paper.

in some sense reveals the true nature of the invertibility of the mapping. Many questions remain open, however. For example, some insight into the cases $w \neq 2^k$ and even n, separately and together, would be desirable.

While the main focus of Thomsen and Rivest was on invertibility, we are more interested in the probability distributions that $g_{w,r_1,\ldots,r_n}$ induce. That is, given an x chosen uniformly at random, what can we say about $g_{w,r_1,\ldots,r_n}(x)$ for different values of w and r_i, how much do the resulting probability distributions differ from the uniform one? To answer this question we will need some more information about the function than an assessment of its invertibility.

From a cryptanalytic perspective, a primitive that exposes a heavily biased probability distribution is prone to distinguishing attacks. This motivates the main goal of this paper, which is to show a few very general results on the probability distributions of Xorrotations. We will then show how to apply our findings toward a cryptanalysis of the eSTREAM final portfolio stream cipher HC-128 [12] designed by Wu.

There have been very few cryptanalytic results on HC-128. In Wu's design paper [12], a distinguisher was constructed based on analysis of the least significant bit of the 32-bit keystream words. The keystream complexity of his distinguisher was about $2^{160.5}$ 32-bit words according to the revised analysis of Stankovski et al. [10]. In that paper they constructed a new distinguisher with a keystream complexity of about $2^{152.5}$, which is the current best for the original HC-128. Other observations and attempts at HC-128 can be found in [2, 4–7].

1.1 Contributions of This Paper

In this paper we present a general treatment of bit mixing using xor and word rotations. We deduce a few simple theoretical results on related probability distributions, results that can be used for cryptanalysis.

We further show how to apply the theory, together with some additional new observations, to produce a new analysis of a variant of the eSTREAM final portfolio stream cipher HC-128. While the keystream complexity of the resulting distinguisher is far from practical ($2^{90.9}$), these are the first results for the given HC-128 variant.

The paper is organized as follows. In Section 2 we first review distinguishers and hypothesis testing before showing some theoretical results on biased probability distributions and bit mixing using xor and word rotations. In Section 3 we apply the theory from the previous section by constructing a new distinguisher for the variant of HC-128 in which addition is replaced with xor. The paper is concluded in Section 4.

2 Biased and RX-Specific Probability Distributions

2.1 Reviewing Distinguishers and Relative Entropy

A distinguisher is a decision mechanism that takes n samples of keystream as input and outputs either "CIPHER" or "RANDOM". The decision mechanism uses the n input samples to perform a hypothesis test to determine which of the two known probability distributions that is more likely, that of the cipher, or the uniform one. The Neyman-Pearson Lemma (see [1]) provides the optimal hypothesis test, which is translated into Definition 1 for independent samples.

For a probability mass function P, we use square bracket notation $P[x]$ here to denote the probability of event x.

Definition 1 (Relative entropy). *The relative entropy between two probability mass functions P_0 and P_1 over the same domain $\mathcal{X}$ is defined as*

$$D\left(P_0 \| P_1\right) = \sum_{x \in \mathcal{X}} P_0[x] \log \frac{P_0[x]}{P_1[x]}. \tag{1}$$

Relative entropy has a couple of aliases in literature; information divergence and Kullback-Leibler distance. In our paper we will sometimes say 'the divergence of P' meaning $D(P\|U)$, where U denotes the corresponding uniform distribution.

The error probabilities for the corresponding hypothesis test reach the point at which they start to decrease exponentially when the number of samples n used in the hypothesis test approaches

$$n \approx \frac{1}{D(P_0\|P_1)}. \tag{2}$$

In this paper, (2) will be used as a measure of the number of samples needed for our distinguisher. This is the same measure that was used in [10].

The time- and keystream complexities of the distinguisher depend on how the observed keystream is used to assemble the samples. In our application, as we will see in Section 3, this assembly is very fast, requiring only four xor operations to build one sample.

2.2 The Divergence of Probabilistically Biased Distributions

We will be using probabilistic equalities in conjunction with the divergence measure D, so we need to determine how the former influence the latter.

Definition 2 (Probabilistically biased distribution). *Let A be any distribution of size 2^w, and let U be the uniform distribution of the same size. A distribution resulting from sampling A with probability p and U with probability $1-p$ is said to be probabilistically biased with parameters (p, A), or (p, A)-biased.*

Theorem 1 (Probabilistic divergence). *The divergence of a (p, A)-biased distribution is $p^2 D(A \| U)$.*

Theorem 1 is proven in the full version of this paper.

2.3 RX-induced Probability Distributions

Consider the function

$$f_{w,r}(x) = x \oplus (x \lll r),$$

where x is a w-bit variable and $\lll$ denotes left rotation with respect to the word length w. For all rotation amounts r in this section we enforce the constraint $0 < r < w$. This construction is often used as a basic mixing component in cryptographic primitives.We take a probability distribution approach here to provide results that are practical for cryptanalysis.

For a distribution to be of use to an analyst, it needs to boast a high divergence. This makes it easily distinguishable from a uniform distribution. In this context, all divergences of magnitude 1 and above are extremely high.

In the following we assume w-bit words, and we number the bit positions from least- to most significant bit 0 through $w-1$. Also, for all rotation amounts r we assume that $0 < r < w$.

Definition 3 (Probability distribution operator E). *A mapping $f : U \longrightarrow V$ is said to* generate *a probability distribution on V (uniformly) in the following way. Starting with an empty array of size $|V|$, let each $x \in U$ contribute probability $2^{-|U|}$ to slot $f(x)$. Summation over all possible domain values produces the probability distribution in question. The probability distribution generated by f is denoted $E(f)$.*

Thus, $f_{w,r}(x)$ and $f_{w,r_1}(x_1) \oplus \ldots \oplus f_{w,r_n}(x_n) = f_{w,r_1,\ldots,r_n}(x_1, \ldots, x_n)$ generate probability distributions $E(f_{w,r})$ and $E(f_{w,r_1,\ldots,r_n})$, respectively.

Definition 4 (r-orbit). *In a w-bit word, the bit positions reachable from bit position i as we apply r-bit rotation again and again – the* orbit *of bit position i under r-bit rotation – is given by the bit set*

$$\{(i + kr) \bmod w \mid k \in \mathbf{N}\},$$

and there are $\gcd(w, r)$ *distinct orbits, each of length* $\frac{w}{\gcd(w,r)}$.

Proposition 1 (Divergence of $E(f_{w,r})$). *The divergence of* $E(f_{w,r})$ *is* $\gcd(w, r)$.

Proof. For every given w-bit output value $y = y_{w-1} \dots y_0$, the equation system

$$f_{w,r}(x) = y$$

has $2^{\gcd(w,r)}$ solutions. That is, restricting the domain and range of $f_{w,r}$ to only one r-orbit, that corresponding equation system has precisely two solutions, and the restricted mapping is consequently 2-to-1. There are $\gcd(w, r)$ disjoint r-orbits, so the entire mapping $f_{w,r}$ is $2^{\gcd(w,r)}$-to-1.

From this it follows that the probability distribution $E(f_{w,r})$ has precisely $2^{w-\gcd(w,r)}$ non-zero probability entries, each being equal to $2^{\gcd(w,r)-w}$ since x is uniformly distributed over the domain. Using (1) we get

$$\begin{aligned} D(E(f_{w,r}) \| U) &= 2^{w-\gcd(w,r)} \left(2^{\gcd(w,r)-w} \log_2 \frac{2^{\gcd(w,r)-w}}{2^{-w}} \right) \\ &= \gcd(w, r). \end{aligned}$$

□

We state a generalized version as Theorem 2, the proof of which can be found in the full version of this paper.

Theorem 2 (Divergence of $E(f_{w,r_1,\dots,r_n})$). *The divergence of* $E(f_{w,r_1,\dots,r_n})$ *is* $\gcd(w, r_1, r_2, \dots, r_n)$.

The probability distribution $E(f_{w,r_1,\cdots,r_n})$ is precisely what we will need for our cryptanalysis of the HC-128 variant, so we will be content with these findings. A deepened RX analysis along the lines of Thomsen [11] and Rivest [9] with further results on $E(g_{w,r_1,\dots,r_n})$ is available in the full version of this paper.

3 Application to HC-128

We now illustrate how Theorems 1 and 2 can be applied in a beautiful way to produce a new distinguisher for a partly linearized version of the stream cipher HC-128. In particular, we show that HC-128 becomes weak if its + operators are replaced with $\oplus$.

Despite the removal of the non-linearity provided by modular addition, it is still *not* easy to construct low-complexity distinguishers for this variant of HC-128. This is because we still have to deal with the S-boxes.

3.1 Notation and Review of HC-128

In this section we give a very brief description of the original HC-128 keystream generation process. HC-128 is defined in [12], from which we adopt and adapt the notation. HC-128 specifies both a key and initialization vector size of 128 bits. Up to 2^{64} bits of keystream can be generated with each key/IV pair. Letting x and y be 32-bit integers, we have binary operators $+, \boxminus, \oplus, ||, \ggg$ and $\lll$ that denote 32-bit addition, subtraction modulo 512, xor, concatenation, and right and left rotation, respectively. The internal state of HC-128 consists of two tables denoted P and Q. Each table contains 512 words of 32 bits each. The keystream is denoted by s and the 32-bit keystream word generated at the i^{th} step is denoted s_i; $s = s_0||s_1||s_2||\ldots$.

Keystream generation proceeds as follows. One table entry is updated and one 32-bit keystream word is generated at each step. One full update of an entire table P or Q takes place during a *session* consisting of 512 consecutive steps. First, table P is updated and table Q is used to provide update values. The roles of tables P and Q are reversed every session.

We will find it convenient to express table entries $P[i]$ as P_i, $P[i \boxminus j]$ as P_{i-j}, and we write P^k_{i-j} for $(P_{i-j} \ggg k)$.

Our analysis is independent of the initialization, so we leave that part out of this description referring to [12].

Probabilistic equalities, equalities that are true with some given minimum probability, are indicated by annotating the equality sign with the corresponding probability.

3.2 A New HC-128 Variant Distinguisher

Table updates in the original HC-128 are performed according to

$$P_i = P_i + (P^{10}_{i-3} \oplus P^{23}_{i-511}) + P^{8}_{i-10}.$$

During the first half session we have 256 table updates with keystream generation

$$s_i = (Q_a + Q_b) \oplus P_i, \tag{3}$$

where $0 \leq i, a \leq 255$ and $256 \leq b \leq 511$. Similarly, the second half session provides 256 table updates with keystream generation

$$s_j = (Q_c + Q_d) \oplus P_j, \tag{4}$$

where $0 \leq c \leq 255$ and $256 \leq j, d \leq 511$. This completes one full update of table P. The subsequent session updates table Q, and for the three first updates we have

$$\begin{aligned} s_k &= (P_l + P_m) \oplus Q_k \\ &= (P_l + P_m) \oplus (Q_e + (Q_f^{10} \oplus Q_g^{23}) + Q_h^8), \end{aligned} \tag{5}$$

with $0 \leq l, e, g \leq 255$ and $256 \leq m, f, h \leq 511$. The P's and Q's in Eq's. (3), (4) and (5) denote lookups into the same tables.

Now consider the HC-128 variant for which all + operators are replaced by $\oplus$. Choosing any one equation triplet (3)(4)(5), we have $i = l$ and $j = m$ (and thus $P_i = P_l$ and $P_j = P_m$) each with probability 2^{-8}. We also have $a = e, c = g$ or $a = g, c = e$ with combined probability 2^{-15} (assume $a = e, c = g$ without loss of generality). We similarly have $b = f, d = h$ or $b = h, d = f$ with combined probability 2^{-15} (assume $b = h, d = f$). Using (3) and (4) linearly together with all the equations (5) gives us 3×256 equation triplets for every 512 keystream words. With probability $\frac{3 \times 256}{512} \times 2^{-46} > 2^{-45.42}$ we therefore have

$$s_i \oplus s_j \oplus s_k \overset{2^{-45.42}}{=} \underbrace{(Q_b \oplus Q_b^8)}_{N_1} \oplus \underbrace{(Q_c \oplus Q_c^{23})}_{N_2} \oplus \underbrace{(Q_d \oplus Q_d^{10})}_{N_3}, \tag{6}$$

where the left-hand side consists of known keystream words only, and N_1, N_2 and N_3 are observations from $E(f_{32,8})$, $E(f_{32,23})$ and $E(f_{32,10})$, respectively. Their combined distribution $E(f_{32,8,23,10})$ has divergence $\gcd(32, 8, 23, 10) = 1$ according to Theorem 2. Eq. (6) shows that we have a $(2^{-45.42}, E(f_{32,8,23,10}))$-biased distribution, which according to Theorem 1 results in a divergence of about $2^{-90.9} \times \gcd(32, 8, 23, 10) = 2^{-90.9}$. This yields a distinguisher requiring roughly $2^{90.9}$ 32-bit keystream words, so it is clear that the + operator plays a vital role in HC-128.

If we use evaluation of the left-hand side of Eq. (6) over all three k-values – four xor operations on 32-bit keystream words – as time unit, we obtain a time complexity of $2^{89.9}$. In absolute terms, this measure is *much* cheaper, a factor of at least 2^{10}, than the cost of an initialization. For comparison, if we were to consider initializations instead, the time complexity of our distinguisher would be less than 2^{80}.

4 Conclusions

We have presented some new and general results on probability distributions related to RX-systems. We have also shown how to apply the new theory to a non-trivial system that uses RX operations in combination with S-boxes. We did this by building a new distinguisher for a partly linearized variant of HC-128. The total time complexity of the new distinguisher is $2^{89.9}$ very simple operations (xor and comparison of 32-bit keystream words) and the distinguisher requires about $2^{90.9}$ 32-bit keystream words.

Acknowledgements

This work was sponsored in part by the Swedish Research Council (Vetenskapsrådet) under grant 621-2006-5249. We also wish to thank Martin Ågren for his helpful comments and discussions.

References

1. T. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley series in Telecommunication. Wiley, 1991.
2. O. Dunkelman. Phorum5: ECRYPT forum, post 'A small observation on HC-128'. Available at *http://www.ecrypt.eu.org/stream/phorum/read.php?1,1143*. Last accessed on January 14, 2011.
3. D. Khovratovich and I. Nikolić. Rotational Cryptanalysis of ARX. In S. Hong and T. Iwata, editors, *Fast Software Encryption 2010*, volume 6147 of *Lecture Notes in Computer Science*, pages 333–346. Springer Berlin / Heidelberg, 2010. *http://dx.doi.org/10.1007/978-3-642-13858-4_19*.
4. A. Kircanski and A. M. Youssef. Differential fault analysis of HC-128. In *Africacrypt 2010*, volume 6055 of *Lecture Notes in Computer Science*, pages 360–377. Springer, 2010.
5. Y. Liu and T. Qin. The key and IV setup of the stream ciphers HC-256 and HC-128. In *International Conference on Networks Security, Wireless Communications and Trusted Computing*, pages 430–433, 2009.
6. S. Maitra, G. Paul, S. Raizada, S. Sen, and R. Sengupta. Some observations on HC-128. *Designs, Codes and Cryptography*, pages 1–15, 2010.
7. G. Paul, S. Maitra, and S. Raizada. A Combinatorial Analysis of HC-128. Cryptology ePrint Archive: Report 2010/387.
8. S. Paul and B. Preneel. Solving systems of differential equations of addition. In C. Boyd and J. González Nieto, editors, *Information Security and Privacy*, volume 3574 of *Lecture Notes in Computer Science*, pages 275–303. Springer Berlin / Heidelberg, 2005. *http://dx.doi.org/10.1007/11506157_7*.
9. R. L. Rivest. The invertibility of the XOR of rotations of a binary word. *International Journal of Computer Mathematics*, 88(2):281–284, January 2011. First published on: December 4, 2010.
10. P. Stankovski, S. Ruj, M. Hell, and T. Johansson. Improved Distinguishers for HC-128. *Designs, Codes and Cryptography*, pages 1–16. *http://dx.doi.org/10.1007/s10623-011-9550-9*.
11. S. S. Thomsen. *Cryptographic hash functions.* PhD thesis, Technical University of Denmark, November 2008.
12. H. Wu. The Stream Cipher HC-128. In *New Stream Cipher Designs*, volume 4986 of *Lecture Notes in Computer Science*, pages 39–47. Springer-Verlag, 2008.