The distinguishing attack on Speck, Simon, Simeck, HIGHT and LEA

Boris Ryabko
Institute of Computational Technologies
Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation and
Novosibirsk State University,
Novosibirsk, Russian Federation

Aleksandr Soskov
Institute of Computational Technologies
Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation

January 7, 2018

Abstract

The purpose of the work is to estimate the resistance of lightweight block ciphers Speck, Simon, Simeck, HIGHT, LEA to a distinguishing attack. (This attack is a form of cryptanalysis on data encrypted by a cipher that allows an attacker to distinguish the encrypted data from random data.) Modern lightweight block ciphers must be designed to be immune to such an attack. It turned out that Speck, Simon, HIGHT and LEA showed a sufficient resistance to the distinguishing attack, but Simeck with 48-bit block size and 96-bit key size was not immune to this attack.

1 Introduction

In recent years a large number of so called lightweight block ciphers have been proposed. They are widely adopted in devices with low computing power. Examples of these highly constrained devices are RFID tags, sensors in wireless sensor network, small internet-enabled appliances from the Internet of Things (IoT), etc. In 2012 an international standard (ISO/IEC 29192) specified the main requirements for lightweight cryptography to provide data confidentiality, authentication, identification, non-repudiation, and key exchange [8].

In this work we analyzed the lightweight block ciphers Speck and Simon [1], Simeck [15], HIGHT [5], LEA [13] regarding to their resistance against

the distinguishing attack. Note, that all ciphers examined in this paper were lightweight block ARX-based ciphers. (The ARX-based ciphers are designed using only modular Addition, Rotation and XOR. In particular, the only source of non-linearity is the modular addition.)

The distinguishing attack is any form of cryptanalysis on data encrypted by a cipher that allows an attacker to distinguish the encrypted data from random data [7]. Thus, one can generate non-random samples of different lengths as an input of the cipher, and the encrypted output text (ciphertext) should not be distinguishable from the random text. If some algorithm can distinguish the output text from random data faster then the direct exhaustive key search, it could be considered as a drawback of the cipher [12].

Our experiments and estimations presented that Speck, Simon, HIGHT and LEA showed a sufficient resistance to the distinguishing attack, but Simeck with 48-bit block size and 96-bit key size was not immune to the distinguishing attack.

The paper is organized as follows. For a more detailed analysed ciphers description the reader is referred to in the section 2. In the section 3 we give the description of the distinguishing attack applied in this work. Also in this section we show the results of calculations obtained for these ciphers. In Section 4 we give theoretical estimations of resistance to the distinguishing attack for Simeck and LEA. In the last section we give a conclusion and some recommendations.

2 Analysed ciphers

In this work we analysed the lightweight block ciphers Speck, Simon, Simeck, HIGHT and LEA.

Simon and Speck are two families of the lightweight block ciphers publicly released by the National Security Agency (NSA, USA) in June 2013. In 2014 it was proposed to include these block ciphers in ISO/IEC 29192 [8].

Simeck is a new family of lightweight block ciphers that combines the properties from both Simon and Speck. The authors of these ciphers consider that Simeck is more compact and efficient block cipher, providing the same level of information security as the Speck and Simon algorithms [15].

HIGHT was proposed in 2006. The first author of HIGHT is also the first author of LEA which was proposed in 2013.

All cipher considered in this work are iterated block ciphers. It means that they transform plaintext divided into blocks of fixed length into identical size blocks of ciphertext by the cyclically repeated invertible function known as the round function, with each iteration referred to as a round [6].

A plaintext to be encrypted (l_0) is processed by the round function for certain number of rounds, and finally the output ciphertext (c_0) is formed, where N denotes the total number of rounds (see fig. 1). The round function uses different keys referred to as round keys, derived from the original key (so-called key schedule). All round functions considered in this work are designed using only modular Addition, Rotation and XOR. The more detailed round

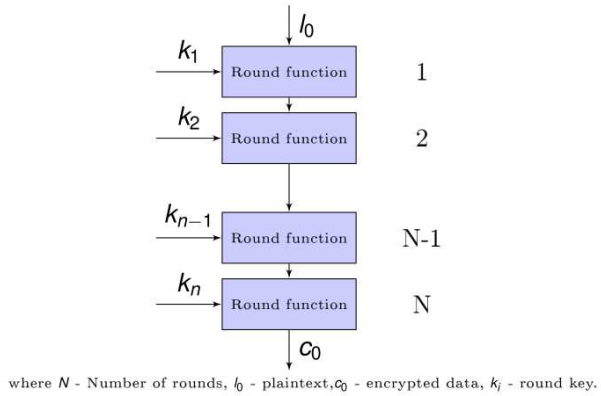

Figure 1: Iterated block cipher scheme

functions definition and key expansions are given in [1], [15], [5], [13]. Block size l_0 is usually equals 32, 48, 64, 96 or 128 bit.

Table 1 lists the different block and key sizes, in bits, for analyzed ciphers and corresponding number of rounds for them.

Speck, Simon and Simeck support a variety of block and key sizes. A block may be 32, 48, 64, 128 bits in size. The corresponding key may be 64, 96, 128, 256 bits in size. The number of rounds depends on the selected parameters. For example, HIGHT has 64-bit block size, 128-bit key size and 32 iterated rounds and LEA has 128-bit block size and 128, 192, or 256-bit key size. The number of rounds also depends on the selected parameters.

3 Experimental results

To determine the resistance of the ciphers to the distinguishing attack we applied the following scheme.

To the input of the cipher we generated "Gray code" sequences, the length of each being equal to the cipher block size (see Appendix B). We note that this input data is very non-random, therefore probability to distinguish encrypted data from random data increases significantly.

At first we encrypted plaintext by the cipher round function for fewer rounds (T) than it was given (the iterated block cipher structure allows us to do it). Then we analysed encrypted text on deviations from randomness by the statistical test "Book Stack". For a more detailed description of the statistical test Book Stack the reader is referred to in the Appendix A.

To distinguish between random and non-random sequences the null hypothesis H_0 that the sequence is random (i.e., its elements are equiprobable and

Table 1: Cipher parameters

Cipher	Block Size(bits)	Key Size(bits)	Rounds
Speck	32	64	22
Speck	48	72, 96	22, 23
Speck	64	96, 128	26, 27
Speck	96	96, 144	28, 29
Speck	128	128, 192, 256	32, 33, 34
Simon	32	64	32
Simon	48	72, 96	36, 36
Simon	64	96, 128	42, 44
Simon	96	96, 144	52, 54
Simon	128	128, 192, 256	68, 69, 72
Simeck	32	64	32
Simeck	48	96	36
Simeck	64	96	44
HIGHT	64	128	32
LEA	128	128, 192, 256	24, 28, 32

independent of each other) must be tested against the alternative hypothesis H_1 that the sequence is not random. A statistical test must decide on whether to accept or reject the null hypothesis. In our experiments the level of significance was 0.001 and computations were made for 100 randomly chosen keys. Thus, if in 100% of cases we observed the hypothesis H_0 is rejected, we concluded that the output sequence was not random.

If we found that the ciphertext was not random we increased the number of round by one and analysed ciphertext encrypted for $T + 1$ rounds. Thus we experimentally determined maximum number of rounds where encrypted data did not look like random bit-sequence. In Table 2 we represented these values for each investigated cipher.

In the column "Parameters" we gave block and key sizes for each cipher in bits. "R_{max}" is maximum number of rounds where encrypted data did not look like random bit-sequence. "R_{full}" is total number of rounds given by the authors of these ciphers. "Length" is the minimal length at which the data encrypted R_{max} times did not look like random bit-sequence.

It turned out that Speck, Simon, HIGHT and LEA showed a sufficient resistance to the distinguishing attack. But for Simeck 48/96 and Simeck 64/128 numbers of rounds where encrypted data did not look like random bit-sequence were large enough. For Simeck 48/96 and Simeck 64/128 these values are equal to 18 and 19 rounds, which are 50% and 43% of the full number of rounds, respectively.

If we increase the computational power, the number of rounds where encrypted data does not look like random will grow and possibly reach the full number of rounds. In order to estimate it we extrapolated the theoretical length

Table 2: The experimental number of rounds where encrypted data did not look like random

Cipher	Parameters	R_{max}	R_{full}	Length
Speck	32/64	5	22	2^{27}
Speck	48/96	6	23	$2^{33.5}$
Speck	64/128	6	27	2^{30}
Speck	96/144	7	29	$2^{33.5}$
Speck	128/256	8	34	2^{35}
Simon	32/64	9	32	2^{27}
Simon	48/96	12	36	$2^{36.5}$
Simon	64/128	12	44	2^{36}
Simon	96/144	17	54	$2^{38.5}$
Simon	128/256	18	72	2^{36}
Simeck	32/64	10	32	2^{29}
Simeck	48/96	18	36	$2^{38.5}$
Simeck	64/128	19	44	2^{36}
HIGHT	64/128	10	32	2^{33}
LEA	128/128	10	24	2^{39}

of the ciphertext required for the successful distinguishing attack on later rounds for Simeck and LEA. These estimations the reader is referred to in the next section.

4 Estimations

For Simeck 48/96, Simeck 64/128 and LEA 128/128 we determined the minimal input sample length at which the encrypted data did not look like random bit-sequence for each round. At first we obtained the sample length experimentally, then we extrapolated the theoretical length required for the successful distinguishing attack on later rounds for these ciphers.

We calculated these extrapolations by the Ordinary Least Squares Method (OLS) using MatLAB software [4].

We note, if we distinguish the output text from random data faster than the direct exhaustive key search, the attack is of interest to cryptologists [12]. Thus, we considered the sample length to be less than 2^{96} for Simeck with 96-bit key and to be less than 2^{128} for Simeck 64/128 and LEA 128/128. These extrapolated results were shown in Table 3.

Our extrapolation for Simeck 48/96 showed that the deviation of the encrypted data from "randomness" could be detected on a full number of rounds with a ciphertext length equal to 2^{96}. It means that Simeck with 48 bit block size and 96 bits key size is not immune to the distinguishing attack.

In case with Simeck 64/128 we could detect the deviation of the encrypted data from "randomness" for the ciphertext length less than 2^{128} for only 34

Table 3: The extrapolated number of rounds where encrypted data did not look like random

Cipher	Block Size(bits)	R_{max}	R_{full}	Length
Simeck	48/96	36	36	2^{96}
Simeck	64/128	34	44	2^{128}
LEA	128/128	16	24	2^{128}

rounds. It means that Simeck 64/128 is resistant to the distinguishing attack.

In case with LEA 128/128 we could detect the deviation of the encrypted data from "randomness" for the ciphertext length less than 2^{128} for only 16 rounds. It means that LEA 128/128 is resistant to the distinguishing attack.

5 Conclusion

We applied the distinguishing attack to modern lightweight block ciphers Speck, Simon, Simeck, HIGHT and LEA using statistical test Book stack. It turned out that Speck, Simon, HIGHT and LEA showed a sufficient resistance to "distinguishing attack", but Simeck with 48-bit block size and 96-bit key size is not immune to this type of attack. We recommend increasing the number of rounds in order to improve the reliability of the Simeck 48/96.

Acknowledgments

The authors wish to thank for the opportunity to perform all the calculations on a supercomputer cluster of Novosibirsk State University.

The authors were supported by the Russian Foundation for Basic Research, Grant no. 15-29-07932.

Appendix A: The statistical test "Book Stack"

A statistical test called "Book Stack" was suggested in [10, 11], see also its description in [9].

Let us briefly describe the book stack test. We have the elements of a sample $X = (x_1, x_2, ..., x_N)$ from the alphabet $A = (a_1, a_2, ..., a_s)$. The book stack test is used to check the hypothesis H_0, that these elements are independent and $P(x_N = a_i) = p^0 = 1/S, n = 1, ..., N; i = 1, ...s$. When the book stack test is applied, all letters from A are ordered from 1 to S and this order (w^n) is changed after observing each letter x_n according to the formula:

$$w^{n+1}(a) = \begin{cases} 1, x_n = a, \\ w^n(a) + 1, w^n(a) < w^n(x_n), \\ w^n(a), w^n(a) > w^n(x_n). \end{cases} \quad (1)$$

This structure is similar to a book stack if we assume that a number coincides with book's position in the stack. The book is extracted from the stack, and after reading, it is put on the top and its number becomes the first. The books that were originally above it are shifted down, and other books remain in their place. If H_0 is not true, then frequent letters from A (as frequently used books) will have relatively small numbers (will spend more time next to the top of the stack). On the other hand, if H_0 is true, the probability to find each letter x_i at each position j is equal to $1/S$.

Before testing the set of all numbers $1, ..., S$ is divided into two non-intersected parts: $A_1 = 1, 2, ...M$ and $A_2 = M + 1, M + 2, ...S$. Then for the sample $(x_1, x_2, ..., x_N)$ we count V_0 - quantity of numbers $w^n(x_n)$ belonging to first part A_1. In other words, V_0 means quantity of hittings in the "upper part" of the "book stack" and $(N - V_0)$ is equal to quantity of hitting in the "lower" part. Then we calculated

$$x^2 = \frac{(V_0 - NP_1)^2}{NP_1} + \frac{((N - V_0) - N(1 - P_1))^2}{N(1 - P_1)} \tag{2}$$

where $P_1 = |A|/S$.

It is known that the x^2 statistic obeys asymptotically the χ^2 (chi-square) distribution (with one degree of freedom in our case, see [14]). The value $\chi^2_{1,1-\alpha}$ is quantile of chi-square distribution with the level of significance equals to $(1 - \alpha)$ with one degree of freedom. Thus, if calculated x^2 is less than critical level $\chi^2_{1,1-\alpha}$, then the hypothesis H_0 is accepted, otherwise rejected.

The present program implementation is due to Alexey Lubkin. It is available at https://github.com/sashasasha-1987/book-stack.

6 Appendix B: Gray code

The reflected binary code, also known as Gray code, is a binary numeral system where two successive values differ in only one bit (binary digit) [2]. This input data is very non-random. We note that it is possible to use Gray code sequence of any length that we need. Besides we can generated a new sequence from the existing by circular shift of all words.

The binary-reflected Gray code list for n bits can be generated recursively from the list $n - 1$ bits by reflecting the list (i.e. listing the entries in reverse order), concatenating the original list with the reversed list, prefixing the entries in the original list with a binary 0, and then prefixing the entries in the reflected list with a binary 1 (see fig. 2).

2-bit list:	00, 01, 11, 10	
Reflected:		10, 11, 01, 00
Prefix old entries with 0:	000, 001, 011, 010,	
Prefix new entries with 1:		110, 111, 101, 100
Concatenated:	000, 001, 011, 010,	110, 111, 101, 100

Figure 2: The Gray code generation for n = 3

References

[1] R. Beaulieu, D. Shors, J. Smith, S. Treatman-Clark, B. Weeks, and L. Wingers. The simon and speck families of lightweight block ciphers. cryptol ogy eprint archive, report 2013/404, 2013.

[2] R. W. Doran. The gray code. *J. UCS*, 13(11):1573–1597, 2007.

[3] F. Hayashi. Econometrics, 2000.

[4] D. Hong, J. Sung, S. Hong, J. Lim, S. Lee, B. Koo, C. Lee, D. Chang, J. Lee, K. Jeong, et al. Hight: A new block cipher suitable for low-resource device. In *CHES*, volume 4249, pages 46–59. Springer, 2006.

[5] P. Junod and A. Canteaut. *Advanced Linear Cryptanalysis of Block and Stream Ciphers (Cryptology and Information Security)*. IOS Press, 2011.

[6] S. Künzli and W. Meier. Distinguishing attack on mag. *ECRYPT Stream Cipher Project Report*, 1:2005, 2005.

[7] K. A. McKay, L. Bassham, M. S. Turan, and N. Mouha. Report on lightweight cryptography. *NIST DRAFT NISTIR*, 8114, 2016.

[8] B. Ryabko and A. Fionov. *Basics of contemporary cryptography for IT practitioners*, World Scientific, 2005.

[9] B. Ryabko and V.Monarev. Using information theory approach to randomness testing. *Journal of Statistical Planning and Inference*, 33(1):95-110, 2005.

[10] B. Y. Ryabko and A. I. Pestunov. "book stack" as a new statistical test for random numbers. *Problems of Information Transmission*, 40(1):66–71, 2004.

[11] B. Schneier. A self-study course in block-cipher cryptanalysis. *Cryptologia*, 24(1):18–33, 2000.

[12] H. Seo, Z. Liu, J. Choi, T. Park, and H. Kim. Compact implementations of lea block cipher for low-end microprocessors. In *International Workshop on Information Security Applications*, pages 28–40. Springer, 2015.

[13] G. Watson et al. Mg kendall, a. stuart, the advanced theory of statistics. *The Annals of Mathematical Statistics*, 42(3):1138–1140, 1971.

[14] G. Yang, B. Zhu, V. Suder, M. D. Aagaard, and G. Gong. The simeck family of lightweight block ciphers. In *International Workshop on Cryptographic Hardware and Embedded Systems*, pages 307–329. Springer, 2015.