χ^2 Attacks on Block-Cipher based Compression Functions

Daniel Santana de Freitas[1], **Jorge Nakahara Jr**

[1]Depto de Informática e Estatística (INE) – Universidade Federal de Santa Catarina (UFSC)
Caixa Postal 476 – 88040-900 – Florianópolis – SC - Brazil

santana@inf.ufsc.br, jorge_nakahara@yahoo.com.br

Abstract. *In this paper, we report on χ^2 analyses of block-cipher based (cryptographic) compression functions. Our aim is not to find collisions nor (second) preimages, but to detect non-random properties that may distinguish a compression function from an ideal primitive such as a random oracle. We study some well-known single-block modes of operation such as Davies-Meyer (DM), Matyas-Meyer-Oseas (MMO) and Miyaguchi-Preneel (MP), and double-block modes such as Hirose's, Tandem-DM, Abreast-DM, Parallel-DM and MDC-2. This paper shows how a weakness (χ^2 correlation) in the underlying block cipher can propagate to the compression function via the mode of operation used in hash constructions. To demonstrate our ideas, we instantiated the block cipher underlying these modes with variable-round RC5, RC6 and ERC6 block ciphers.*

Keywords: χ^2 cryptanalysis, block-cipher-based (cryptographic) compression functions, single- and double-block-length modes of operation.

1. Introduction

Hash and compression functions are pervasive cryptographic primitives used for privacy and authentication purposes in environments as diverse as computer networks, sensor networks and mobile devices [Kaufman et al. 2002, Stallings 2003]. There are many security properties that compression and hash functions have to satisfy depending on the application environment. In this paper, we apply χ^2 analysis techniques to block-cipher based compression and hash functions in order to detect some nonrandom behavior that demonstrates that they are not ideal primitives. Our aim is not to find collisions nor (second) preimages [Menezes et al. 1997], but statistical correlations between the input and the output of the compression or hash functions. We recall that NIST requested in [NIST 2007b] that the SHA-3 candidates should behave as close as possible to random oracles [NIST 2007a], which is as expected of an ideal cryptographic primitive.

We selected the following block ciphers to instantiate the permutation mapping(s), denoted E, inside the compression functions under analysis (Fig. 1):

- the RC6 block cipher [Rivest et al. 1998] has a Feistel Network structure and was a finalist candidate to the AES competition [NIST 2001] aimed at selecting the successor of the DES [NIST 1993] cipher. RC6 actually stands for a family of ciphers, and is more appropriately designated as RC6-$w/r/b$, where w stands for the word size in bits (RC6 is a word-oriented cipher in the sense that it operates on neatly partitioned w-bit words), r is the number of rounds and b is the key length in bytes. For the AES competition, $w = 32$ (a text block consists of four w-bit

words), $r = 20$ and three key sizes: 128 bits ($b = 16$), 192 bits ($b = 24$) and 256 bits ($b = 32$). This version is called simply RC6.

- the RC5 block cipher [Rivest 1994] has a Feistel Network structure, and is the predecessor of RC6. More formally, RC5 stands for a family of ciphers, and is designated RC5-$w/r/b$, with parameters w, r and b having the same meaning as in RC6 and value ranges: $w \in \{16, 32, 64\}$, $0 \leq k \leq 255$, $0 \leq r \leq 255$. A text block in RC5 consists of two w-bit words. The original parameter values are $w = 32$ (leading to a 64-bit block), $r = 12$ (nowadays, 20 rounds are suggested) and $b = 16$ (or 128-bit key).
- the ERC6 block cipher [Ragab et al. 2001] has a Feistel Network structure and is a large-block generalization of RC6, thus the name Extended RC6. Formally, ERC6 stands for a family of ciphers, and is parameterized as ERC6-$w/r/b$, with the same meaning as for RC6. A text block in ERC6 consists of eight w-bit words. Nominal values of these parameters are not available, but in [Ragab et al. 2001] one can find the values $w = 32$, $r = b = 16$ used for performance evaluation in software.

This choice of block ciphers leads to hash digests (or chaining variables) of 64, 128, 256 and 512 bits for the single- and double-block length hash modes. Our attacks do not depend on related-keys nor on the key schedule algorithms. Further details about it, for each key size, can be found in [Rivest et al. 1998, Rivest 1994, Ragab et al. 2001].

Our choice of RC5/RC6/ERC6 was motivated by well-known results of χ^2 attacks on reduced-round versions of each cipher [Isogai et al. 2002, Knudsen 2002, Jr. et al. 2009]. These results indicate that not only a large number of rounds can be distinguished from a random permutation without exhausting the codebook, but also that the χ^2 correlation patterns are *iterative*, which is a fundamental feature for leveraging the attack to a large number of rounds when these ciphers are used as building blocks in compression and hash function constructions.

This paper is organized as follows: Sect. 2 summarizes the contributions of this paper. Sect. 3 briefly explains a χ^2 analysis. Sect. 4 describes attacks on the compression function in DM mode. Sect. 5 describes attacks on compression functions in MMO and MP modes. Sect. 6 presents an attack on a double-block-length mode of operation called Tandem-DM. Sect. 7 presents an attack on the Abreast-DM mode. Sect. 8 presents an attack on the Parallel-DM mode. Sect. 9 presents an attack on the MDC-2 mode. Sect. 10 concludes this paper.

2. Contributions

The contributions of this paper include

- a concrete application of χ^2 technique [Knudsen and Meier 2000] to compression functions; in particular we aim at block-cipher based hash function constructions.
- to demonstrate our attacks, we instantiated the block cipher(s) inside the hashing modes of operation with variable-round versions of RC5 [Rivest 1994], RC6 [Rivest et al. 1998] and ERC6 [Ragab et al. 2001], for which there are well-known χ^2 correlations for a large number of rounds. We require the target block ciphers to be susceptible to χ^2 (distinguishing) attacks, so that we can extend this correlation further to the compression function, via the mode of operation, and eventually

to the hash function. Note that the fact that up to 44-round ERC6 (a large number of rounds) is susceptible to χ^2 attacks already indicates that these primitives are not appropriate building blocks in compression function constructions.

- further, *we exploit the fact that the correlations patterns are iterative*, that is, they have the same form both at the input and at the output of the block cipher, so that we can bypass the feedforward and feedback of chaining variables commonly found in modes of operation, and thus extend the distinguishing capability of χ^2 technique beyond the block cipher framework.
- we perform only *distinguishing attacks* since there are no secret keys in compression/hash function settings. On one hand, this fact implies that the attack complexities are lower since no keys are recovered/guessed. On the other hand, the fact that the key is not secret means that it is under the control of the adversary. So, in our attacks, we can freely choose the key to be weak concerning the χ^2 attacks, and thus, cover a large number of rounds of the underlying block ciphers. Even though χ^2 attacks can be performed under a known-plaintext setting [Shimoyama et al. 2001, Miyaji et al. 2002], we have to choose some message blocks and chaining variables to satisfy the attack requirements. Thus, all our attacks operate under a chosen-message-or-chaining-variable (CMCV) setting.
- we analyse single-block modes of operation such as Davies-Meyer (DM), Matyas-Meyer-Oseas (MMO) and Miyaguchi-Preneel (MP) [Menezes et al. 1997]. This is a subset of a large series of modes of operation originally described by Preneel *et al.* in [Preneel et al. 1994]. We also analysed double-block modes such as Hirose's [Hirose 2006], Tandem-DM, Abreast-DM [Lai and Massey 1993], Parallel-DM and MDC-2 [Brachtl et al. 1990], [Meyer and Schilling 1988].
- our attacks use the χ^2 technique to demonstrate non-random properties of the underlying block cipher. We describe how a particular weakness, a χ^2 correlation, can propagate from the block cipher to the compression function via the mode of operation. Therefore, we *do not* aim at traditional attacks such as collisions or (second) preimage, but our findings are relevant in applications where hash functions are expected to behave as random oracles such as (pseudo-)random number generators, which is required by NIST for the ongoing SHA-3 competition [NIST 2007b, NIST 2007a]. While most of the traditional analysis of hash functions use differential cryptanalysis (aimed at finding collisions), our approach uses the χ^2 technique for the analysis of compression functions in order to uncover weaknesses and non-random behavior which prove that these functions are not ideal cryptographic primitives.

3. A Brief Overview of χ^2 Correlations

We recall from [Knudsen and Meier 2000] a brief description of the χ^2 goodness-of-fit test to distinguish a random source with unknown probability distribution p_X from a random source with uniform distribution p_U. This χ^2 test will be used to detect correlations between specific input and output bits of r-round ciphers. Let $X = x_0, x_1, \ldots, x_{n-1}$ be identically and independently distributed random variables taking values in the set $\{a_0, a_1, \ldots, a_{m-1}\}$ with unkown probability distribution. The χ^2 test is used to decide if an observation $x_0, x_1, \ldots, x_{n-1}$ is consistent with the hypothesis $\Pr(x = a_j) = p(j)$ for $0 \leq j < m$, where $p_X = \{p(j)\}$ is a discrete probability distribution on a set of m

elements. Let $N_{a_j}(X)$ denote the number of times the observation X takes on the value a_j. Then, $\sum_j N_{a_j}(X) = n$. The χ^2 statistic is the random variable defined by

$$\chi^2 = \sum_{j=0}^{m-1} \frac{(N_{a_j}(X) - np(j))^2}{np(j)}.$$

For the uniform distribution p_U, the χ^2 statistic is just $\frac{m}{n}\sum_j (N_{a_j}(X) - \frac{n}{m})^2$. In a χ^2 test, the observed χ^2 statistic is compared to $\chi^2_{a,m-1}$, the threshold for the χ^2 test with $m-1$ degrees of freedom and with significance level a.

In this paper, we exploit χ^2 correlations as a distinguishing tool to detect non-random behavior of block ciphers inside a compression function in modes of operation such as Matyas-Meyer-Oseas (MMO), Davies-Meyer (DM) and Miyaguchi-Preneel (MP) [Menezes et al. 1997]. Further, we look to extend such correlations to the full mode of operation, and eventually to the hash function as well. Since there is no key involved in compression or hash functions, all attacks are of the distinguish-from-random type.

In [Knudsen and Meier 2000], Knudsen and Meier showed χ^2 correlations can be used to distinguish 15-round RC6 from a random permutation (and up to 17 rounds for weak keys). These correlations were derived from 2-round iterative linear relations called Type-I approximations, having the form

$$(A \cdot e_t) \oplus (C \cdot e_s) = (A' \cdot e_u) \oplus (C' \cdot e_v), \tag{1}$$

where A and C are the first and third w-bit input words, that is, a plaintext block has the form $A\|B\|C\|D$, while A' and C' are the first and third w-bit words two rounds later, that is, the cipher state after two rounds has the form $A'\|B'\|C'\|D'$; e_t denotes a w-bit mask with a single bit 1 in the t-th least significant bit position. If t, s, u and v are nonzero but less than five, there is a nonzero bias which depends on the values of t, s, u, v. The idea is to fix each of the $\text{lsb}_5(A)$ and $\text{lsb}_5(C)$ and collect statistics of the ten bits: $\text{lsb}_5(A')$ and $\text{lsb}_5(C')$. That means an iterative distinguisher that can be concatenated every two rounds. Note that the χ^2 analysis sets the least significant $\log_2 w$ bits of A and C to zero. This is a consequent of previous analysis of RC5 in [Borst et al. 1999].

RC6 is well-known for the extensive use of data-dependent rotations. It was mentioned in [Knudsen and Meier 2000] that for some weak keys that generate zero rotation amounts, χ^2 correlations could be detected up to 16-round RC6. Since we control the key input, which means the message blocks m_i or chaining variables H_{i-1} and H_i, we can afford to choose them so that the key is weak. Also for this same reason, we can include the pre-whitening layer with subkeys. But, we omit the post-whitening layer since our distinguishers do not cover the full nominal number of rounds. Therefore, following [Knudsen and Meier 2000], we assume that with about 2^{118} chosen plaintexts, we can detect χ^2 correlation across 16-round RC6 under weak keys.

In [Miyaji et al. 2002], Miyaji *et al.* described improvements to χ^2 attacks applied to RC5-32 ($w = 32$), for up to 10 rounds. The analysis approach is similar to equation (1). Let a plaintext block be denoted $A\|B$ and the output after h rounds be $A'\|B'$. The attack setting in reduced-round RC5 starts by setting $\text{lsb}_5(B)$ to zero and computing χ^2 correlation in $\text{lsb}_5(B')$. The idea is similar to (1). For an even number of rounds:

$$B \cdot e_t = B' \cdot e_u, \tag{2}$$

where e_t denotes a w-bit mask with a single bit 1 in the t-th least significant bit position. The best χ^2 distinguishing attack reported in [Miyaji et al. 2002] reaches 10 rounds and requires $2^{57.87}$ chosen plaintexts and equivalent 10-round computations.

In [Shimoyama et al. 2001], Shimoyama *et al.* described χ^2 analysis of RC5-32/10 using $2^{57.87}$ chosen ciphertexts and equivalent number of 10-round computations. This attack has high success rate and uses a similar approach to (2). For weak keys that lead to small or zero rotation values, χ^2 correlations can be detected further up to 12 rounds.

In [Isogai et al. 2002], Isogai *et al.* described attacks on RC5-64, ie. with $w = 64$ up to 16 rounds. This is an indication that increasing the word size helps improve the attack to more rounds since χ^2 correlations can be detected over a larger set of $\log_2 w$ bits of some input and output words of r-round RC5-64. Actually, as demonstrated by attacks on RC5, RC6, ERC6 and further on MRC6 [El-Fishawy et al. 2004], in this order, it seems that increasing the number of words in a text block also helps improve the χ^2 attacks to a larger number of rounds.

In [Jr. et al. 2009], Nakahara *et al.* described χ^2 distinguishing attacks on up to 44-round ERC6 using 2^{233} chosen plaintexts. There is no mention of attack improvements due to weak keys/subkeys in [Jr. et al. 2009]. So, we apply our analysis assuming the key is fixed but arbitrary. The correlation pattern for ERC6 has a similar form to (1) but a text block has eight words, denoted $A\|B\|C\|D\|E\|F\|G\|H$. For an even number of rounds, and following a Type-I approximation, the correlation pattern for ERC6 is

$$A \cdot e_{t_1} \oplus C \cdot e_{t_2} \oplus E \cdot e_{t_3} \oplus G \cdot e_{t_4} = A' \cdot e_{t_5} \oplus C' \cdot e_{t_6} \oplus E' \cdot e_{t_7} \oplus G' \cdot e_{t_8}. \tag{3}$$

There are many other ciphers that could be attacked, but they are all variants of RC5/RC6, weakened or extended versions such as RC6-64 [Knudsen 2002] and MRC6 in [El-Fishawy et al. 2004]. We limit our analyses to these 3 ciphers: RC5, RC6 and ERC6, since they provide enough freedom of choice of block and key sizes.

4. Attacks on the DM mode

The DM mode is depicted in Fig. 1(b). The i-th chaining value is computed as

$$H_i = H_{i-1} \oplus E_{m_i}(H_{i-1}), \tag{4}$$

where $H_0 = IV$ is the initial value, m_i is the i-th message block and $E_x(y)$ is a block cipher with key x and plaintext y. In all settings, we assume the key size to E can be adapted to fit the message or chaining variable length.

Let us consider RC6-32/16/16 as E for an initial attack on a compression function in DM mode since equation (1) is iterative, that is, the very same words are used both at the input and the output. In particular, let $m_i, H_{i-1} \in \mathbb{Z}_2^{128}$ be the input and $H_i \in \mathbb{Z}_2^{128}$ be the output of a compression function in DM mode. The attack proceeds as follows: we fix m_i appropriately as a weak key so that we our attack reaches 16 rounds of E. We choose H_{i-1} as input to E according to (1), that is, *with the five least significant bits of its first and third words set to zero* so that a χ^2 correlation can be detected in the same words at $E_{m_i}(H_{i-1})$.

Despite the feedforward of H_{i-1} in DM mode in (4), this correlation can be *extended* to H_i because of the way H_{i-1} was constructed, that is, the zero bits in H_{i-1} do not affect the correlation in H_i and the correlation pattern appears at the same place in both H_i and $E_{m_i}(H_{i-1})$. Following [Knudsen and Meier 2000], by appropriately choosing H_{i-1} and keeping m_i fixed, we can detect a biased distribution in H_i, and thus, distinguish a compression function in DM mode instantiated by RC6-32/16/16 from a random mapping using 2^{118} chosen chaining variable values and equivalent effort. Memory is negligible since we do not need to store the messages, but only the χ^2 statistics. Note that even if the feedforward of H_{i-1} was combined via modular addition instead of xor, the attack would still hold because the least significant five bits of the 1st and 3rd words of H_{i-1} were purposefully set to zero.

Suppose we set $i = 2$ in the previous paragraph. If we vary m_1 over all 2^{128} possible text blocks, with H_0 fixed, then $H_1 = E_{m_1}(H_0) \oplus H_0$ would behave as a random function. According to [Rijmen et al. 1997], varying m_1 this way would cover about $\frac{1}{e} \cdot 2^{128} = 2^{126.56}$ of the codebook. Suppose now that we keep m_2 fixed as key input to the next compression function instance, then $H_2 = E_{m_2}(H_1) \oplus H_1$. If m_2 is the last message block then it should contain padding and the total message length (following MD strengthening). But, we cannot control H_1 to have zero bits in the appropriate positions for the χ^2 attack on the underlying block cipher. Thus, our attack is restricted to the compression function only.

Now, suppose we instantiate E with RC5-32/10/8 in DM mode. We apply the iterative pattern in (2). This time, $m_i, H_{i-1} \in \mathbb{Z}_2^{64}$ are the input and $H_i \in \mathbb{Z}_2^{64}$ is the output of a compression function in DM mode. The attack proceeds as follows: we fix m_i as a key so that E behaves as a permutation. We choose H_{i-1} as input to E according to (2), that is, with the five least significant bits of its second word only set to zero so that a χ^2 correlation can be detected in the same words at $E_{m_i}(H_{i-1})$. Despite the feedforward of H_{i-1} in (4), this correlation can be leveraged to H_i because of the way H_{i-1} was constructed, that is, the zero bits in H_{i-1} do not affect the correlation in H_i and the correlation pattern appears at the same place in both H_i and $E_{m_i}(H_{i-1})$. Following [Miyaji et al. 2002], by appropriately choosing H_{i-1} and keeping m_i fixed, we can detect a biased distribution in H_i, and thus, distinguish a compression function in DM mode instantiated by RC5-32/10/8 from a random mapping using $2^{57.87}$ chosen messages and equivalent effort.

Repeating the same reasoning for ERC6-32/44/32 in DM mode, and applying the pattern (4). This time $m_i, H_{i-1} \in \mathbb{Z}_2^{256}$ are the input and $H_i \in \mathbb{Z}_2^{256}$ are the output of the compression function. The attack proceeds as follows: we fix m_i as key so that E behaves as a permutation. We choose H_{i-1} as input to E so that the five least significant bits of four words, according to (4), are set to zero so that a χ^2 correlation can be detected in the same words at $E_{m_i}(H_{i-1})$, despite the feedforward of H_{i-1} in (4), as explained in the previous paragraphs. Following [Jr. et al. 2009], by appropriately choosing H_{i-1} and keeping m_i fixed, we can detect a biased distribution in H_i, and thus, distinguish a compression function in DM mode instantiated by ERC6-32/44/32 from a random mapping using 2^{233} chosen messages and equivalent effort.

A reasoning why so far we could not extend the attack to a hash function in DM mode is the following: suppose a message $M = m_1 \| m_2$ where m_1 is variable while H_0

is fixed. It means that $H_1 = H_0 \oplus E_{m_1}(H_0)$ behaves as a random function. According to [Rijmen et al. 1997], if m_1 runs over all 2^{128} values, then about $\frac{1}{e} \cdot 2^{128} \approx 2^{126.56}$ distinct values appear in H_1. We keep m_2 fixed since it might contain (eventual) padding and the length of M, due to the Merkle-Damgaard (MD) strengthening. Thus, $E_{m_2}(H_1)$ behaves as a permutation. If we could filter out only those values of H_1 which have zero bits in the five least significant positions of its first and third words, then we could detect a χ^2 correlation in $H_2 = H_1 \oplus E_{m_2}(H_1)$, which would be the hash digest. Out of the $2^{126.56}$ inputs, 10 bits would have to be fixed. That means only $2^{116.56}$ would be available, which is not enough. We need 2^{118} texts.

5. Distinguishing attacks for the MMO and MP modes

In MMO mode, the i-th chaining variable is computed as

$$H_i = m_i \oplus E_{H_{i-1}}(m_i), \tag{5}$$

with $H_0 = IV$, as depicted in Fig. 1(a). Let us consider RC6-32/16/16 as E in MMO mode. Let $m_i, H_{i-1} \in \mathbb{Z}_2^{128}$ be the input and $H_i \in \mathbb{Z}_2^{128}$ be the output of the compression function. The attack proceeds as follows: we choose a fixed (weak) value for H_{i-1} so that it is a (weak) key to E. We choose m_i as input to E according to (1), that is, with the five least significant bits of its first and third words set to zero so that a χ^2 correlation can be detected in the same words at $E_{H_{i-1}}(m_i)$. Due to the feedforward of m_i, this χ^2 correlation can be extended to H_i because of (5). Following [Knudsen and Meier 2000], by appropriately choosing m_i and keeping H_{i-1} fixed, we can detect a biased distribution in H_i, and thus, distinguish a compression function MMO mode instantiated by RC6-32/16/16 from a random mapping using 2^{118} chosen messages and equivalent effort. Memory is negligible since we do not need to store the messages, but only the χ^2 statistics.

Now, suppose we instantiate E with RC5-32/10/8 in MMO mode. The attack would proceed similarly to the one for RC6-32/16/16, but m_i, H_{i-1}, $H_i \in \mathbb{Z}_2^{64}$ and the complexity becomes $2^{57.87}$ chosen messages and equivalent effort.

Analogously, if we instantiate E with ERC6-32/44/32 in MMO mode. Once again, the attack would proceed similarly to the one for RC6, but m_i, H_{i-1}, $H_i \in \mathbb{Z}_2^{256}$ and the complexity would become 2^{233} chosen messages and equivalent effort.

In MP mode, the i-th chaining variable is computed as

$$H_i = m_i \oplus H_{i-1} \oplus E_{H_{i-1}}(m_i), \tag{6}$$

with $H_0 = IV$, as depicted in Fig. 1(c). We consider RC6-32/16/16 instantiating E in MP mode. The attack proceeds analogous to the one in the previous paragraph. The MP and MMO modes are quite similar, except for a feedforward of H_{i-1}. We keep H_{i-1} fixed as a (weak) key input. Thus, we can simple remove it from H_i before checking the χ^2 correlation, by just xoring H_i with H_{i-1} in (6). The attack complexity remains the same as in the MMO case.

Now, if we instantiate E with RC5-32/10/8 in MP mode, the attack on the compression function would proceed similarly to the one for RC6-32/16/16, but m_i, H_{i-1}, $H_i \in \mathbb{Z}_2^{64}$ and the complexity becomes $2^{57.87}$ chosen messages and equivalent effort.

Analogously, we instantiate E with ERC6-32/44/32 in MP mode. Once again, the attack on the compression function would proceed similarly to the one for RC6-32/16/16, but m_i, H_{i-1}, $H_i \in \mathbb{Z}_2^{256}$ and the complexity would become 2^{233} chosen messages and equivalent effort.

6. Distinguishing attacks on the Tandem-DM mode

In [Lai and Massey 1993], Lai and Massey proposed a double-block length hash mode called Tandem-DM that uses two instances of an n-bit block, $2n$-bit key cipher E. A compression function instance is depicted in Fig. 1(e). The initial values are H_0^1 and H_0^2.

The attack proceeds as follows: we instantiate E with RC6-32/16/32, message blocks $m_i \in \mathbb{Z}_2^{128}$ and chaining variables H_i^1, $H_i^2 \in \mathbb{Z}_2^{128}$. We keep H_{i-1}^1 and m_i fixed as key input, so that $E_{H_{i-1}^1 \| m_i}(.)$ behaves as a permutation. We vary H_{i-1}^2 over 2^{118} distinct messages, with the least significant five bits of its first and third words set to zero. We shall detect χ^2 correlation at $E_{H_{i-1}^1 \| m_i}(H_{i-1}^2)$ with high probability. Because of our choice of H_{i-1}^2, its feedforward will still allow to detect correlation at $H_i^2 = H_{i-1}^2 \oplus E_{H_{i-1}^1 \| m_i}(H_{i-1}^2)$.

Due to the feedback of $E_{H_{i-1}^1 \| m_i}(H_{i-1}^2)$ as part of the key input to the other E instance, our attack is restricted to the compression function (and only to one E instance). The attack complexity is 2^{118} chosen messages and equivalent computation effort.

Now, if we instantiate E with RC5-32/10/16 (we have to adjust the key size to fit the double-sized key input) in Tandem-DM mode. The attack on the compression function proceeds similarly to the one in the previous paragraph, but m_i, H_{i-1}, $H_i \in \mathbb{Z}_2^{64}$ and the complexity becomes $2^{57.87}$ chosen messages and equivalent effort.

Analogously, if we instantiate E with ERC6-32/44/64 in Tandem-DM mode. Once again, the attack on the compression function would proceed similarly to the previous ones, but m_i, H_{i-1}, $H_i \in \mathbb{Z}_2^{256}$ and the complexity would become 2^{233} chosen messages and equivalent effort.

7. Distinguishing attacks on Abreast-DM mode

In [Lai and Massey 1993], Lai and Massey proposed a double-block length hash mode called Abreast-DM. A compression function instance is depicted in Fig. 1(f).

The attack proceeds as follows: we instantiate E with RC6-32/16/32. We keep H_{i-1}^1 and m_i fixed as key inputs, so that $E_{H_{i-1}^1 \| m_i}(.)$ behaves as a permutation. We vary H_{i-1}^2 over 2^{118} distinct messages, all of which have the pattern (1). We shall detect χ^2 correlation at $E_{H_{i-1}^1 \| m_i}(H_{i-1}^2)$ with high probability. Because of our choice of H_{i-1}^2, its feedforward will still allow to detect correlation at H_i^2.

Due to the feedforward of H_{i-1}^2 as part of the key input to the other E instance, our attack is restricted to the compression function (and only to one E instance). In the original definition of Abreast-DM, the E instance whose input is H_{i-1}^1 is negated (bitwise NOT). For our attacks, it does not matter since we do not depend on this E instance. We use only the other E instance[1]. Anyway, if necessary, we could compensate for this bitwise NOT by xoring with '1' bits the specific positions where we look for χ^2

[1]Thus, as long as the E instance we attack contains RC6-32/16/32, the other E instance could be any block cipher whatsoever.

correlation. We can afford to do it since we attack the compression function. Thus, for simplicity, we omit the bitwise NOT in Fig. 1(f) and in the attack description.

The final attack complexity is 2^{118} chosen messages and equivalent computational effort.

Now, if we instantiate E with RC5-32/10/16 in Abreast-DM mode. The attack on the compression function would proceed similarly to the one in the previous paragraph, but $m_i, H_{i-1}, H_i \in \mathbb{Z}_2^{64}$ and the complexity becomes $2^{57.87}$ chosen messages and equivalent effort.

Analogously, if we instantiate E with ERC6-32/44/64 in Abreast-DM mode. Once again, the attack on the compression function would proceed similarly to the previous ones, but $m_i, H_{i-1}, H_i \in \mathbb{Z}_2^{256}$ and the complexity would become 2^{233} chosen messages and equivalent effort.

8. Distinguishing attacks on the Parallel-DM mode

The Parallel-DM is a double-block length hash mode of operation designed by Hohl *et al.* in [Hohl et al. 1993]. A compression function instance is depicted in Fig. 1(g).

Let us consider RC6-32/16/32 as E in Parallel-DM mode. We attack the compression function only. Let $m_i^1, m_i^2 \in \mathbb{Z}_2^{128}$ be fixed values, as key inputs to the E instance whose input is $H_{i-1}^1 \in \mathbb{Z}_2^{128}$. The attack proceeds as follows: we choose a weak value for $m_i^1 \oplus m_i^2$ so that it is a (weak) key to E, which behaves as a permutation. We choose $H_{i-1}^1 \oplus m_i^1$ as input to E so that the least significant five bits of its first and third words are zero, while the remaining bits vary over 2^{118} values. This means that H_{i-1}^1 shall be chosen carefully to compensate for the xor with a fixed m_i^1 before being input to E. This way, a χ^2 correlation can be detected in the same words at $E_{m_i^1 \oplus m_i^2}(m_i^1 \oplus H_{i-1}^1)$. Due to the feedforward of $H_{i-1}^1 \oplus m_i$, this χ^2 correlation can be extended to H_i^1 because of the way H_{i-1}^1 was constructed. Following [Knudsen and Meier 2000], by appropriately choosing m_i^1, m_i^2 and H_{i-1}^1, we can detect a biased distribution in H_i^1, and thus, distinguish a compression function Parallel-DM mode instantiated by RC6-32/16/32 from a random mapping using 2^{118} chosen messages and equivalent effort. Memory is negligible since we do not need to store the messages, but only the χ^2 statistics.

Now, if we instantiate E with RC5-32/10/8 in Parallel-DM mode the attack on the compression function would proceed similarly to the one in the previous paragraph, but $m_i, H_{i-1}, H_i \in \mathbb{Z}_2^{64}$ and the complexity becomes $2^{57.87}$ chosen messages and equivalent effort.

Analogously, we instantiate E with ERC6-32/44/32 in Parallel-DM mode. Once again, the attack on the compression function would proceed similarly to the previous ones, but $m_i, H_{i-1}, H_i \in \mathbb{Z}_2^{256}$ and the complexity would become 2^{233} chosen messages and equivalent effort.

9. Distinguishing attacks on MDC-2 mode

The MDC-2 mode is a double-block length hash mode described by Bracthl *et al.* in [Brachtl et al. 1990] and in [Meyer and Schilling 1988]. It is depicted in Fig. 1(h). According to [Menezes et al. 1997], there is a fixed word swap after the feedforward of m_i,

before the output is assigned to H_i^1 and H_i^2, but this swap does not affect our attack. Thus, we omit this swap to simplify the analysis.

Let us consider RC6-32/16/16 as E in MDC-2 mode. We attack the compression function only. Let $H_{i-1}^1, H_{i-1}^2 \in \mathbb{Z}_2^{128}$ be fixed values, as key inputs to the E instances, and $m_i \in \mathbb{Z}_2^{128}$. The attack proceeds as follows: we choose a weak value for H_{i-1}^1 and H_{i-1}^2 so that they are (weak) keys to the two E's, that behave as permutations. We choose m_i so that the least significant five bits of its first and third words are zero, while the remaining bits vary over 2^{118} values. This way, a χ^2 correlation can be detected in the same words at $E_{H_{i-1}^1}(m_i)$ and at $E_{H_{i-1}^2}(m_i)$. Due to the feedforward of m_i in both E instances, this χ^2 correlation can be detected in H_i^1 and H_i^2 as well, because of the way m_i was constructed. Following [Knudsen and Meier 2000], we can detect a biased distribution in both H_i^1 and H_i^2 and thus, distinguish a compression function MDC-2 mode instantiated by RC6-32/16/16 from a random mapping using 2^{118} chosen messages and equivalent effort. Memory is negligible since we do not need to store the messages, but only the χ^2 statistics.

Now, we instantiate E with RC5-32/10/8 in MDC-2 mode. The attack on the compression function would proceed similarly to the one in the previous paragraph, but $m_i, H_{i-1}, H_i \in \mathbb{Z}_2^{64}$ and the complexity becomes $2^{57.87}$ chosen messages and equivalent effort.

Analogously, if we instantiate E with ERC6-32/44/32 in MDC-2 mode, the attack on the compression function would proceed similarly to the previous ones, but m_i, H_{i-1}, $H_i \in \mathbb{Z}_2^{256}$ and the complexity would become 2^{233} chosen messages and equivalent effort.

10. Conclusions

This paper presented a χ^2 analysis of block-cipher based compression functions. Our attacks included both single-block-length modes such as Matyas-Meyer-Oseas (MMO), Davies-Meyer (DM) and Miyaguchi-Preneel (MP) [Menezes et al. 1997], and double-block-length hash modes such as Tandem-DM, Hirose's, Abreast-DM, Parallel-DM and MDC-2. Attack complexities are listed in Table 1. Memory complexity is negligible, since we do not need to store the messages, and since we perform distinguishing attacks only. There is no point in key-recovery attacks since there is no secret value involved.

We describe how χ^2 correlations can be extended from the underlying block ciphers to the compression function, in several modes of operation, so that these primitives do not behave ideally, but can be distinguished from a random mapping in a chosen-text setting.

In our analyses, we used variable-round versions of ERC6, RC6 and RC5 ciphers to instantiate the block cipher(s) underlying all modes of operation studied. Our choice of block ciphers is based on the fact that these versions are susceptible to χ^2 cryptanalysis. Nonetheless, they serve as a proof-of-concept. Our point is to make clear that weaknesses such as χ^2 correlations can propagate beyond the block cipher to a compression function in some hashing mode of operation such as DM, MMO and MP, allowing one to distinguish them from random mappings. Unlike other attacks, we do not exploit this weakness to find collisions nor (second) preimages. Rather, we search for other properties (χ^2 correlations) that have been overlooked and that may indicate that these primitives are

Table 1. χ^2 attack complexities on compression functions in several modes of operation. Memory cost is negligible.

Data (CMCV)	Time	Modes	Block Cipher
2^{233}	2^{233}	DM, MMO, MP, Parallel-DM, MDC-2	ERC6-32/44/32
2^{233}	2^{233}	Tandem-DM, Abreast-DM	ERC6-32/44/64
2^{118}	2^{118}	DM, MMO, MP, Parallel-DM, MDC-2	RC6-32/16/16
2^{118}	2^{118}	Tandem-DM, Abreast-DM	RC6-32/16/32
$2^{57.87}$	$2^{57.87}$	DM, MMO, MP, Parallel-DM, MDC-2	RC5-32/10/8
$2^{57.87}$	$2^{57.87}$	Tandem-DM, Abreast-DM	RC5-32/10/16

CMCV: Chosen Message or Chaining Variable

not ideal candidates for applications such as pseudorandom number generators and PRP's [NIST 2007a]. We believe these properties are relevant and shall not be underestimated.

Our results were possible due to the fact that the correlation patterns for RC5, RC6 and ERC6 are iterative, that is, they have the same form for both the input and the output of the block cipher. This simple fact allowed us to bypass the feedback and feedforward of values commonly used in modes of operation.

So far, our analyses apply to the compression functions only. We leave as an open problem how to extend our χ^2 correlations attacks to the full hash function.

References

Borst, J., Preneel, B., and Vandewalle, J. (1999). Linear cryptanalysis of RC5 and RC6. In *Fast Software Encryption (FSE)*, LNCS 1636, pages 16–30. Springer.

Brachtl, B., Coppersmith, D., Hyden, M., Jr., S. M., Meyer, C., Oseas, J., Pilpel, S., and Schilling, M. (1990). Data authentication using modification detection codes based on a public one-way encryption function. US Patent 4908861.

El-Fishawy, N., Danaf, T., and Zaid, O. (2004). A modification of RC6 block cipher algorithm for data security (MRC6). In *International Conference on Electrical, Electronic and Computer Engineering*, pages 222–226.

Hirose, S. (2006). Some plausible constructions of double-block length hash functions. In Robshaw, M., editor, *Fast Software Encryption Workshop, FSE 2006*, LNCS 4047, pages 210–225. Springer.

Hohl, W., Lai, X., Meier, T., and Waldvogel, C. (1993). Security of iterated hash functions based on block ciphers. In D.R.Stinson, editor, *Adv. in Cryptology, Crypto 1993*, LNCS 773, pages 379–390. Springer.

Isogai, N., Miyaji, A., and Nonaka, M. (2002). Cryptanalysis of RC5-64 with improved correlation attack. In *SCIS 2002, The 2002 Symposium on Cryptography and Information Security*, pages 657–662. The Institute of Electronics, Information and Communications Engineers.

Jr., J. N., Sekar, G., de Freitas, D. S., Chiann, C., de Souza, R. H., and Preneel, B. (2009). A new approach to χ^2 cryptanalysis of block ciphers. In P.Samarati, editor, *Information Security Conference (ISC)*, LNCS 5735, pages 1–16. Springer.

Kaufman, C., Perlman, R., and Speciner, M. (2002). *Network Security: PRIVATE Communication in a PUBLIC World*. Prentice-Hall.

Knudsen, L. (2002). Correlations in RC6 on 256-bit blocks. NESSIE technical report nes/doc/uib/wp5/022/1, Univ. of Bergen.

Knudsen, L. and Meier, W. (2000). Correlations in reduced round variants of RC6. In B.Schneier, editor, *Fast Software Encryption, FSE 2000, 7th International Workshop*, LNCS 1978, pages 94–108. Springer.

Lai, X. and Massey, J. (1993). Hash function based on block ciphers. In *Adv. in Cryptology, Eurocrypt 1992*, LNCS 658, pages 55–70. Springer.

Menezes, A., van Oorschot, P., and Vanstone, S. (1997). *Handbook of Applied Cryptography*. CRC Press.

Meyer, C. and Schilling, M. (1988). Secure program load with manipulation detection code. In *Proceedings 6th Worldwide Congress on Computer and Communications Security and Protection (SECURICOM 1988)*, pages 111–130.

Miyaji, A., Nonaka, M., and Takii, Y. (2002). Improved correlation attack on RC5. IEICE Trans. on Fundamentals, vol. E85-A, no. 1, 44–57.

NIST (1993). FIPS: Data encryption standard. Federal Information Processing Standards Publication 46-2, supersedes FIPS PUB 46-1.

NIST (2001). FIPS: Advanced encryption standard. Federal Information Processing Standards Publication 197.

NIST (2007a). Announcing request for candidate algorithm nominations for a new cryptographic hash algorithm (SHA-3) family. Federal Register, vol. 72, no. 212, Nov.

NIST (2007b). Cryptographic hash algorithm competition. available at http://csrc.nist.gov/groups/ST/hash/sha-3/index.html.

Preneel, B., Govaerts, R., and Vandewalle, J. (1994). Hash functions based on block ciphers: a synthetic approach. In *Adv. in Cryptology, Crypto 1993*, LNCS 773, pages 368–378. Springer.

Ragab, A., Ismail, N., and Allah, O. (2001). Enhancements and implementation of RC6 block cipher for data security. IEEE TENCOM.

Rijmen, V., Preneel, B., and Win, E. D. (1997). On weaknesses of non-surjective round functions. *Design, Codes and Cryptography*, 12(3):253–266.

Rivest, R. (1994). The RC5 encryption algorithm. In *Proceedings 2nd International Workshop on Fast Software Encryption (FSE)*, LNCS, pages 86–96. Springer.

Rivest, R., Robshaw, M., Sidney, R., and Yin, Y. (1998). The RC6 block cipher. http://www.rsa.com/rsalabs.

Shimoyama, T., Takeuchi, K., and Hayakawa, J. (2001). Correlation attack to the block cipher RC5 and the simplified variants of RC6. Advanced Encryption Standard (AES3) Conference, Jun.

Stallings, W. (2003). *Cryptography and Network Security: Principles and Practice*. Prentice Hall.

A. Modes of Operation Schematics

Fig. 1 illustrates several modes of operation discussed in this paper. E stands for a block cipher; the triangle in the side of the square holding E indicates the key input; c is a constant.

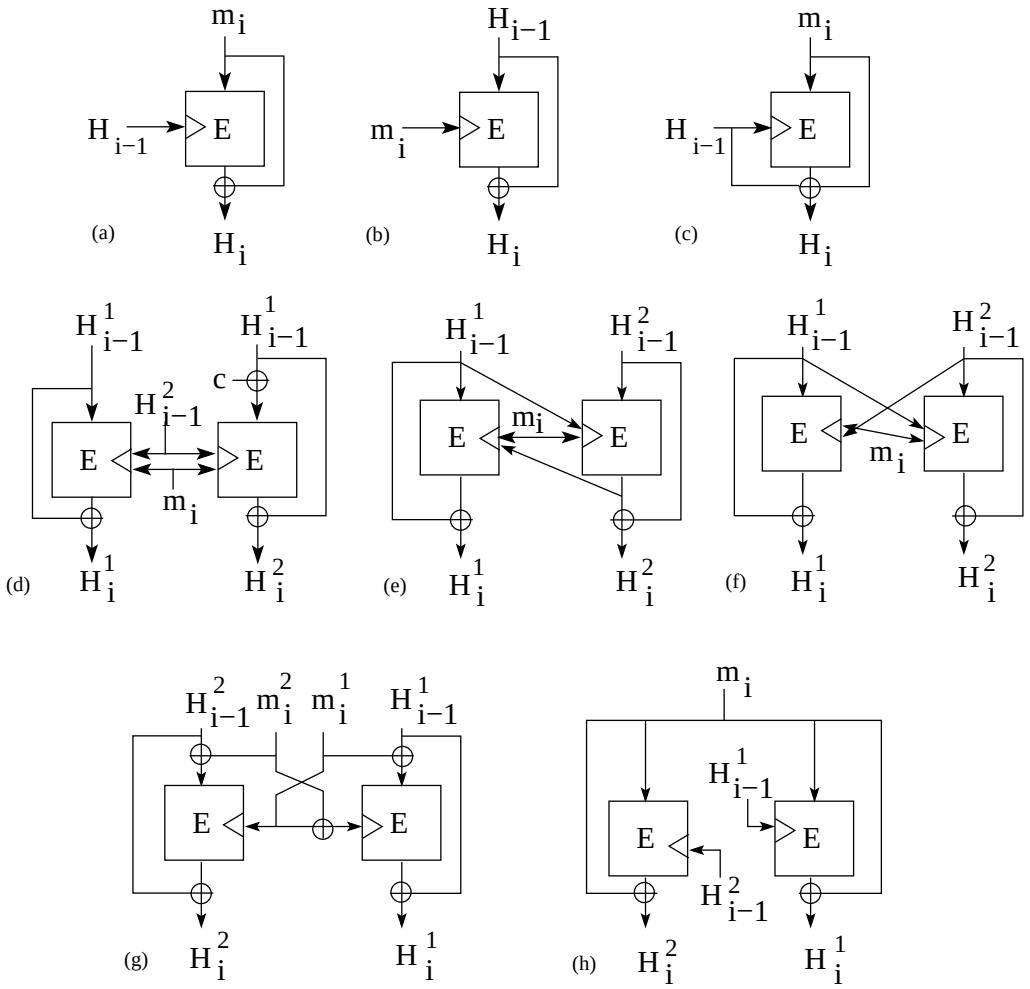

Figure 1. Modes of operation: (a) Matyas-Meyer-Oseas (MMO); (b) Davies-Meyer (DM); (c) Miyaguchi-Preneel (MP); (d) Hirose's; (e) Tandem-DM; (f) Abreast-DM; (g) Parallel-DM; (h) MDC-2

