Non-random Properties of Compression and Hash Functions using Linear Cryptanalysis

Daniel Santana de Freitas[1] and Jorge Nakahara Jr.[2]
[1]*Dept. of Computer Science, Federal University of Santa Catarina, Santa Catarina, Brazil*
[2]*Dept. d'Informatique, Université Libre de Bruxelles, Brussels, Belgium*

Keywords: Linear Analysis, Block-Cipher-Based Hash Functions, Tandem-DM, Abreast-DM, Parallel-DM.

Abstract: We report on linear analyses of block-cipher based compression and hash functions. Our aim is not to find collisions nor (second) preimages, but to detect non-random properties that may distinguish a compression or hash function from an ideal primitive (random oracle). We study 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, Parallel-DM and Abreast-DM. This paper points out weaknesses coming from the feedforward operation used in these hash modes. We use an inside-out approach: we show how a weakness (linear relation) in the underlying block cipher can propagate to the compression function and eventually to the whole hash function. To demonstrate our ideas, we instantiate the block cipher underlying these modes with 21-round PRESENT, the full 16-round DES and 9-round Serpent. For instance, in DM-PRESENT-80 mode, we can distinguish the hash function from an ideal primitive with 2^{64} hash computations.

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 (C.Kaufman et al., 2002). In this paper, we apply the linear cryptanalysis (LC) technique to block-cipher-based compression and hash functions in order to detect nonrandom behaviours that demonstrate some instances are not ideal primitives. Our aim is not to find collisions nor (second) preimages, but linear relationships between the input message and the output chaining variable or hash digest. In (NIST, 2007), NIST requested that candidate hash functions should behave as close as possible to random oracles.

To instantiate the block cipher(s) inside the compression functions, we chose:

- PRESENT, a Substitution-Permutation-Network (SPN) design, operating on 64-bit text blocks, iterating 31 rounds and using keys of 80 or 128 bits (A.Bogdanov et al., 2007).
- Data Encryption Standard (DES) (FIPS, 1993) is a 64-bit Feistel cipher parameterized by a 56-bit key and iterating 16 rounds.
- Serpent is a 128-bit SPN cipher, with keys of 128, 192 and 256 bits, and iterating 32 rounds (Anderson et al., 1998).

Our attacks are independent of the key schedule algorithms. The reason for selecting these block ciphers is because there are well-known linear relations covering a large number of rounds with high bias.

Linear cryptanalysis (LC) was developed by M. Matsui (M.Matsui, 1994) and aimed at the DES (FIPS, 1993; Matsui, 1994) and FEAL ciphers. LC exploits *linear approximations* which stand for a linear combination of bits of the plaintext, ciphertext and key bits holding with high, nonzero bias. In this paper, we exploit linear relations in the underlying block cipher(s) as a distinguishing tool to detect non-random behavior of compression functions in modes of operation such as Matyas-Meyer-Oseas (MMO), Davies-Meyer (DM), Miyaguchi-Preneel (MP) (Menezes et al., 1997), Hirose's (S.Hirose, 2006), Tandem-DM, Abreast-DM (Lai and Massey, 1993) and Parallel-DM. Moreover, *we look to leverage these linear relations to the full mode of operation and eventually to the hash function as well.* Since there is no key involved in the compression and hash functions, all attacks are of the distinguish-from-random type.

This paper is organized as follows: Sect. 2 summarizes the contributions of this paper; Sect. 3 describes the modes of operation under analyses; Sect. 4

Santana de Freitas D. and Nakahara Jr J..
Non-random Properties of Compression and Hash Functions using Linear Cryptanalysis.
DOI: 10.5220/0004475204710477
In *Proceedings of the 10th International Conference on Security and Cryptography* (SECRYPT-2013), pages 471-477
ISBN: 978-989-8565-73-0

describes attacks to compression and hash functions; Sect. 5 concludes this paper.

2 CONTRIBUTIONS

The contributions of this paper include

- a concrete application of linear cryptanalysis (LC) (M.Matsui, 1994) to block-cipher based compression and hash functions. We analyse both single-block modes of operation such as DM, MMO and MP, and double-block length modes such as Hirose's, Tandem-DM, Parallel-DM and Abreast-DM.
- our attacks demonstrate non-random properties of compression/hash functions. We use an *inside-out approach*: we describe how a weakness (linear relation) in the block cipher can propagate to the compression function via the mode of operation and eventually to the entire hash function.
- in the case of DM mode, we were able to attack the full hash function (see Table 1). However, our attacks do not contradict the results of Damgård (I.B.Damgård, 1989), since our aim is to detect nonrandom behavior of the hash function, while Damgard was concerned with collision resistance.
- in attacks on hash functions, such as DM and Parallel-DM, we use iterative linear relations with low Hamming-weight bit masks. If the bits exploited in the positions specified by the mask are not truncated in the hash digest then our attacks still hold. This fact indicates that *truncating the hash digest*, a common practice to adapt the digest size to different applications, is not enough to avoid our attacks.
- our findings are relevant in applications where hash functions are expected to behave as random mappings such as pseudorandom number generators, which is required by NIST for the SHA-3 competition (NIST, 2007). While most of the traditional analysis of hash functions use differential cryptanalysis (DC), aiming at finding collisions, our approach uses LC in order to uncover weaknesses and non-random behavior which prove that the compression or hash function are not ideal primitives.

3 HASHING MODES

Let $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher parameterized by a k-bit key and operating on n-bit blocks. The g mapping transforms its input to the appropriate key size if necessary, otherwise, g is omitted; m_i is the i-th message block and $H_0, H_0^1, H_0^2 \in \{0,1\}^n$ are the initial values. The i-th chaining variables H_i or (H_i^1, H_i^2) are computed as follows

- Davies-Meier (DM):

$$H_i = H_{i-1} \oplus E_{g(m_i)}(H_{i-1}). \quad (1)$$

- Matyas-Meyer-Oseas (MMO):

$$H_i = m_i \oplus E_{g(H_{i-1})}(m_i). \quad (2)$$

- Miyaguchi-Preneel (MP):

$$H_i = m_i \oplus H_{i-1} \oplus E_{g(H_{i-1})}(m_i). \quad (3)$$

- Hirose's mode:

$$H_i^1 = H_{i-1}^1 \oplus E_{g(H_{i-1}^2 \| m_i)}(H_{i-1}^1), H_i^2 = H_{i-1}^1 \oplus E_{g(H_{i-1}^2 \| m_i)}(c \oplus H_{i-1}^1), \quad (4)$$

  where $c \in \{0,1\}^n$ is a nonzero constant.
- Tandem-DM, a double-block length hash mode, uses two instances of an n-bit block, $2n$-bit key cipher E:

$$H_i^1 = H_{i-1}^1 \oplus E_{m_i \| E(H_{i-1}^2)}(H_{i-1}^1), H_i^2 = H_{i-1}^2 \oplus E_{H_{i-1}^1 \| m_i}(H_{i-1}^2). \quad (5)$$

- Parallel-DM:

$$H_i^1 = H_{i-1}^1 \oplus m_i^1 \oplus E_{g(m_i^1 \oplus m_i^2)}(H_{i-1}^1 \oplus m_i^1), \quad (6)$$

$$H_i^2 = H_{i-1}^2 \oplus m_i^2 \oplus E_{g(m_i^1 \oplus m_i^2)}(H_{i-1}^2 \oplus m_i^2). \quad (7)$$

- Abreast-DM:

$$H_i^1 = H_{i-1}^1 \oplus E_{m_i \| H_{i-1}^2}(H_{i-1}^1), H_i^2 = H_{i-1}^2 \oplus E_{H_{i-1}^1 \| m_i}(H_{i-1}^2). \quad (8)$$

4 ATTACKS

Attacks on DM Mode. The designers of the block cipher PRESENT (A.Bogdanov et al., 2007) defined two modes of operation for compression functions in (A.Bogdanov et al., 2008): DM-PRESENT-80 (80-bit key) a single-block mode giving a 64-bit hash digest; and H-PRESENT-128 (128-bit key) for 128-bit digest, which is a double-block-length hash mode. The DM-PRESENT-80 mode is the Davies-Meier mode adapted to the PRESENT cipher with an 80-bit key. See (1) where E is the PRESENT cipher, $n = 64$ and $k = 80$. The H-PRESENT-128 mode is Hirose's mode

adapted to PRESENT with a 128-bit key. See (4), where E is the PRESENT cipher, $n = 64$, $k = 128$ and g maps its input to a 128-bit key.

Concerning the linear relations and linear hulls for PRESENT, we exploit the analysis in (Nakahara.Jr et al., 2009) and from which we adopt the same terminology. We denote by Γ a 64-bit mask, unless stated otherwise. The dot (or inner) product between two bit strings is denoted by a $\cdot$, for instance, $a \cdot b = \bigoplus_{i=0}^{63} a_i \cdot b_i$, for 64-bit strings a and b. The bit mask ΓP indicates the input linear relation, and ΓC will denote the output linear relation for a given cipher. There are no explicit details about the use of Merkle-Damgård (MD) strengthening (I.B.Damgård, 1989; Merkle, 1989) or otherwise in the hash functions derived from PRESENT. For the attacks described in this paper, we assume the usual MD strengthening: padding (a single bit '1' followed by as many '0' bits as necessary) and the message length as part of the last message block.

According to (Nakahara.Jr et al., 2009), the bit mask $\Gamma = \texttt{0000000000200000}_{\chi}$ achieves a good trade-off in linear attacks because of: (i) its low Hamming Weight, (ii) the high bias across a large number of rounds (e.g. 21 rounds) because of small number of active S-boxes, and (iii) the fact that it is *iterative* that is, the same bit mask is used for both the input and output text block. Thus, the linear relation for the underlying cipher E using PRESENT that we analyse has the general form

$$x \cdot \Gamma \oplus E_k(x) \cdot \Gamma = k \cdot \Gamma_1, \tag{9}$$

where Γ_1 is a fixed bit mask for the key. Thus, we use the bit masks $\Gamma = \Gamma P = \Gamma C$ for attacking DM-PRESENT-80.

The attack on the full hash function proceeds as follows: we instantiate the block cipher E with 21-round PRESENT in DM-PRESENT-80. This attack is possible because the linear relation is *iterative*, which allows to make the linear relation depend only on the hash digest (details below). We assume the message M to be hashed has two blocks, $M = m_1 \| m_2$. Since the message blocks m_i are input as key into PRESENT in DM-PRESENT-80, we vary the 64-bit m_1 thus, H_1 changes accordingly because $H_1 = H_0 \oplus E_{g(m_1)}(H_0)$, but we keep m_2 fixed for all messages M. Since we change the key for a fixed plaintext, $E_{g(m_1)}(H_0)$ does not behave as a permutation but as a random function. According to (V.Rijmen et al., 1997), using all 2^{64} values of m_1 we expect to obtain about $2^{64}/e \approx 2^{62.56}$ distinct values from $E_{g(m_1)}(H_0)$, where $e \approx 2.718$ is the base of natural logarithms. According to (Nakahara.Jr et al., 2009), the bias of the linear relation is $2^{-30.11}$, and this amount of plaintext still allows to achieve a high success rate attack. Note that m_2 contains $|M|$ and some padding due to the MD strengthening. Therefore, H_1 as plaintext input to E will vary, but since m_2 is fixed, the E instance for the second compression function will behave as a permutation.

We apply the linear approximation (9) to the second instance of E. Notice that the linear relation covering 21-round PRESENT is $H_1 \cdot \Gamma \oplus E_{g(m_2)}(H_1) \cdot \Gamma = m_2 \cdot \Gamma_1$, where Γ_1 is a fixed bitmask corresponding to the key, which is $g(m_2)$. Notice that in DM-PRESENT-80, there is a feedforward of the H_1 value. Since the same mask Γ is used in both the input and output of E, the linear relation for the full compression function, the DM-PRESENT-80 mode, becomes (using the conventional rules of propagating bit masks across xor and branching structures)

$$H_2 \cdot \Gamma = m_2 \cdot \Gamma_1, \tag{10}$$

that is, the dependence on H_1 disappears because $H_2 = H_1 \oplus E_{g(m_2)}(H_1)$. We do not need to know Γ_1 nor m_2. Since both values are fixed, $m_2 \cdot \Gamma_1$ is fixed as well. Since only the parity of the linear relation matters, (10) can be simplified to

$$H_2 \cdot \Gamma = 0. \tag{11}$$

This setting is similar to a ciphertext-only attack, because the mask Γ and the feedforward of the DM mode makes the linear relation depend on H_2 only. Since M has two blocks, H_2 is the hash digest. Therefore, using 2^{64} messages M we can distinguish DM-PRESENT-80 from a random mapping by analysing the parity of a single bit from the hash digest alone. For a random mapping, the relation (11) might hold with a much lower bias (much closer to zero). Note that the mask Γ is very special: it has low Hamming Weight and the bits that participate in the linear approximation are clustered. Therefore, even if H_2 were truncated the attack would still apply as long as the parity bit indicated by Γ is not truncated. Note that even though the linear attack requires only known plaintext, we have to choose different m_1 blocks to force H_1 to change, while keeping m_2 fixed. Therefore, this attack requires chosen plaintexts/messages/chaining variables.

The hash digest in this case is only 64 bits, which is not large enough to provide a significant level of security, either concerning collision, (second) preimage or other relevant property. Even in a lightweight setting, this hash digest size might not be enough. Nonetheless, *this attack is a proof-of-concept: it demonstrates how a weakness (linear relation) in the underlying block cipher can propagate to the mode of operation (compression function) and further to the hash function in DM mode, by detecting a bias in the hash digest alone.*

Distinguishing Attack using the Full 16-round DES. Let us instantiate the block cipher inside the DM mode with the full DES (FIPS, 1993). In particular, for our distinguishing setting, we employ the 16-round linear relation described in the annex of (M.Matsui, 1994)[p.397], from which we adopt the terminology for bit numbering. The linear relation of covering the full DES has plaintext mask $\Gamma P = (\Gamma P_L, \Gamma P_R) = ([7,18,24],[12,16])$, ciphertext bitmask $\Gamma C = (\Gamma C_L, \Gamma C_R) = ([15],[7,18,24,29,27,28,30,31])$ and bias $1.49 \cdot 2^{-24}$. Here ΓP_L, ΓP_R, ΓC_L, ΓC_R are each 32-bit masks. Note that $\Gamma P \neq \Gamma C$, that is, the relation is not iterative. This fact will limit our attack to the compression function only. The attack proceeds as follows: consider the full 16-round DES as E in DM mode. We assume that g removes parity bits to adjust the 64-bit message to 56 bits. According to (M.Matsui, 1994), the bias of this linear relation is $1.49 \cdot 2^{-24}$, which leads to $8 \cdot (1.49 \cdot 2^{-24})^{-2} = 2^{49.84}$ messages for a high success rate attack. We assume the message block m_i to be fixed (as the key) for all $2^{49.84}$ input blocks H_{i-1}. So, $E_{g(m_i)}(H_{i-1})$ behaves as a permutation. The linear relation around the block cipher is $H_{i-1} \cdot \Gamma P \oplus E_{g(m_i)}(H_{i-1}) \cdot \Gamma C = g(m_i) \cdot \Gamma_1$, where Γ_1 is the mask for the key. The exact value of Γ_1 is $K_1[19,23] \oplus K_3[22] \oplus K_4[44] \oplus K_5[22] \oplus K_7[22] \oplus K_8[44] \oplus K_9[22] \oplus K_{11}[22] \oplus K_{12}[44] \oplus K_{13}[22] \oplus K_{15}[22] \oplus K_{16}[42,43,45,46]$, where K_i denotes the i-th round subkey. In any case, $g(m_i) \cdot \Gamma_1$ is a fixed parity bit. Propagating the masks to the DM-mode, we obtain $H_i \cdot \Gamma C = H_{i-1} \cdot (\Gamma P \oplus \Gamma C)$, where $H_i = H_{i-1} \oplus E_{g(m_i)}(H_{i-1})$. In summary, we have $H_i \cdot \Gamma C = H_{i-1} \cdot (\Gamma P \oplus \Gamma C)$. Thus, by analysing the input and output of DM mode with 16-round DES as E, we can distinguish the compression function from a random oracle using $2^{49.84}$ messages. Note that unlike in Sect. 4, this time the linear relation surrounding E is not iterative. For this reason we cannot propagate it backwards to attack the hash function. On the positive side, we cover the full DES cipher instead of a reduced-round cipher.

Attacks in MMO and MP Modes The MMO mode for PRESENT follows (2) with $n = 64$, $k = 80$ and we call it MMO-PRESENT-80. Let $g : \mathbb{Z}_2^{64} \rightarrow \mathbb{Z}_2^{80}$ be an injective, deterministic mapping that transforms a 64-bit block into an 80-bit key. The exact g transformation is not important. For our attack purposes, if x is fixed then $g(x)$ is fixed as well, and vice-versa. A similar attack to that on DM-PRESENT-80 can be applied to MMO-PRESENT-80. This attack proceeds as follows: consider 21-round PRESENT as E in MMO-PRESENT-80. According to (Nakahara.Jr et al., 2009), the bias is $2^{-30.11}$, which leads to $2^{63.22}$ messages for a high success rate distinguishing attack. We keep H_{i-1} fixed so that $g(H_{i-1})$ is a fixed key. We vary m_i over $2^{63.22}$ messages and apply the linear approximation (9) to E. Notice that the linear relation covering 21-round PRESENT is $m_i \cdot \Gamma \oplus E_{g(H_{i-1})}(m_i) \cdot \Gamma = g(H_{i-1}) \cdot \Gamma_1$, for some fixed bitmask Γ_1 associated to the key $g(H_{i-1})$. In (2) there is a feedforward of m_i. Since the same mask Γ is used in both the input and output of E, the linear relation for the compression function becomes $H_i \cdot \Gamma = g(H_{i-1}) \cdot \Gamma_1$, that is, the dependence on m_i disappears. We do not need to know Γ_1 nor $g(H_{i-1})$ because both are fixed, thus $g(H_{i-1}) \cdot \Gamma_1$ is a fixed bit parity, and the relation reduces to $H_i \cdot \Gamma = 0$. Again, this setting is similar to a ciphertext-only attack, because the mask Γ and the feedforward of m_i makes the linear relation depend on the output H_i only. Since we vary m_i, this message block (which would contain padding and the length of M) is not fixed and thus, the attack applies only to the compression function. Therefore, using $2^{63.22}$ messages we can distinguish the compression function in MMO-PRESENT-80 from a random mapping.

A similar attack can be adapted to the MP mode (3), with $n = 64$ and $k = 80$ and it corresponds to MP-PRESENT-80. The mapping $g : \mathbb{Z}_2^{64} \rightarrow \mathbb{Z}_2^{80}$ transforms a 64-bit string into an 80-bit key. The exact g transformation is not important. For our attack purposes, if x is fixed, then $g(x)$ is a fixed value as well, and vice-versa. Our attack proceeds as follows: consider 21-round PRESENT as E in MP-PRESENT-80. According to (Nakahara.Jr et al., 2009), the bias for the linear relation using $\Gamma = \texttt{0000000000200000}_\texttt{x}$ is $2^{-30.11}$, which leads to $2^{63.22}$ messages for a high success rate attack. We assume H_{i-1} to be fixed and input as key into PRESENT in MP-PRESENT-80. So, $g(H_{i-1})$ is also fixed. We vary the 64-bit m_i over $2^{63.22}$ messages and apply the linear approximation (9) to E. The linear relation covering 21-round PRESENT is $m_i \cdot \Gamma \oplus E_{g(H_{i-1})}(m_i) \cdot \Gamma = g(H_{i-1}) \cdot \Gamma_1$, for some bitmask Γ_1 associated with the key $g(H_{i-1})$. In (3) there is a feedforward of both m_i and H_{i-1}. Since the same mask Γ is used in both the input and output of E, the linear relation for the compression function becomes $H_i \cdot \Gamma = H_{i-1} \cdot (\Gamma_1 \oplus \Gamma)$, i.e. the dependence on m_i disappears. We do not need to know Γ_1 nor $g(H_1)$ because both values are fixed. Since we vary m_i, the last message block (which might contain padding and the length of M) cannot be a fixed value, and this attack applies only to the compression function.

Attack on H-PRESENT-128. For H-PRESENT-128 we have a double chaining variable: $(H_i^1, H_i^2) \in \mathbb{Z}_2^{64} \times \mathbb{Z}_2^{64}$. The attack proceeds as follows: suppose

21-round PRESENT as E in H-PRESENT-128. No g transformation for the key input is needed in this case since the key is 128 bits. Our attack is restricted to the compression function. According to (Nakahara.Jr et al., 2009), the bias of the linear relation with mask $\Gamma = \texttt{0000000000200000}_\chi$ is $2^{-30.11}$, which leads to $2^{63.22}$ messages for a high success rate attack. We assume both H^2_{i-1} and m_i are fixed values because they are input as keys into the two instances of PRESENT in H-PRESENT-128, which will behave as permutations. As for H^1_{i-1}, we use $2^{63.22}$ distinct values as plaintext input to both E instances. We can apply the linear approximation with bit mask Γ to either instance of E. For one of them, the linear relation covering the 21-round PRESENT in E is

$$H^1_{i-1} \cdot \Gamma \oplus E_{H^2_{i-1}\|m_i}(H^1_{i-1}) \cdot \Gamma = H^2_{i-1} \cdot \Gamma_1 \oplus m_i \cdot \Gamma_2, \quad (12)$$

where Γ_1 and Γ_2 are the bit masks for the key $H^2_{i-1}\|m_i$. The right-hand-side of (12) is fixed since H^2_{i-1}, m_i and the masks are fixed. So (12) can be simplified to $H^1_i \cdot \Gamma = 0$, whose format is due to the feedforward of H^1_{i-1} value: $H^1_i = H^1_{i-1} \oplus E_{H^2_{i-1}\|m_i}(H^1_{i-1})$. The analogous linear relation for the second E instance is $(H^1_{i-1} \oplus c) \cdot \Gamma \oplus E_{H^2_{i-1}\|m_i}(H^1_{i-1}) \cdot \Gamma = H^2_{i-1} \cdot \Gamma_1 \oplus m_i \cdot \Gamma_2$. Thus, we can detect bias in both chaining variables H^1_i and H^2_i. Using $2^{63.22}$ messages we can distinguish the compression function of H-PRESENT-128 from an ideal mapping. For a random mapping, the relation $H^1_i \cdot \Gamma = 0$ would hold with a much lower bias (much closer to zero), so that $2^{63.33}$ messages will not be enough to detect any bias. Because of the use of H^2_{i-1} as key, our attack is restricted to the compression function only.

Attack on Tandem-DM Mode. We apply a linear attack to the *compression function in Tandem-DM mode* (5) with message blocks $m_i \in \mathbb{Z}_2^{64}$, and H^1_i, $H^2_i \in \mathbb{Z}_2^{64}$. We use PRESENT with 128-bit key so that there is no need for a transformation g prior to the key input. The attack proceeds as follows: suppose 21-round PRESENT in both instances of E in Tandem-DM. According to (Nakahara.Jr et al., 2009), the bias of the linear relation with mask $\Gamma = \texttt{0000000000200000}_\chi$ is $2^{-30.11}$, which leads to $2^{63.22}$ messages for a high success rate attack. We assume m_i and H^1_{i-1} are fixed as key to the one of the E instances, while H^2_{i-1} varies over $2^{63.22}$ distinct values. We apply the linear approximation with bit mask Γ to both the input and the output of the compression function labeled by H^2_{i-1} and H^2_i. We obtain the linear relation:

$$H^2_{i-1} \cdot \Gamma \oplus E_{H^1_{i-1}\|m_i}(H^2_{i-1}) \cdot \Gamma = (H^1_{i-1}\|m_i) \cdot \Gamma_1, \quad (13)$$

where Γ_1 is the bit masks for the key $H^1_{i-1}\|m_i$. The right-hand-side of (13) is a fixed parity bit since H^1_{i-1}, m_i and Γ_1 are fixed values. Due to the feedforward of H^2_{i-1}, (13) can be simplified to $H^2_i \cdot \Gamma = 0$, that is, there is no more dependence on H^2_{i-1} nor on H^1_{i-1} since $H^2_i = H^2_{i-1} \oplus E_{H^1_{i-1}\|m_i}(H^2_{i-1})$. The distinguishing attack depends only on H^2_i. Therefore, using $2^{63.22}$ messages we can distinguish this compression function in Tandem-DM from an ideal mapping. Due to feedback of $E_{H^1_{i-1}\|m_i}(H^2_{i-1})$ as part of the key input in $E_{m_i\|E_{H^1_{i-1}\|m_i}(H^2_{i-1})}$ in the second E instance, we do not analyse the full chaining value (H^1_i, H^2_i). But analysing H^2_i is enough to attack the compression function.

Distinguishing Attack using Reduced-round Serpent We use the results on linear cryptanalysis on 9-round Serpent with 256-bit key described in (E.Biham et al., 2002) to attack a compression function in Tandem-DM mode. The attack proceeds similarly to that on PRESENT, except that: (i) the bias is 2^{-52} and therefore, $8 \cdot (2^{-52})^{-2} = 2^{107}$ values H^2_{i-1} are required for a high success rate attack; (ii) the bit masks for 9-round Serpent are not iterative. We call the input and output bit masks simply $\Gamma P = [14, 24, 25, 26, 44, 45, 46, 48, 49, 60, 62, 63, 74, 84, 86, 87, 86, 87, 88, 89, 90, 100, 103, 114]$ and $\Gamma C = [15, 35, 52, 75, 80, 81, 82, 93, 121, 122]$. These bits were derived from the bit-slicing representation of Serpent and following the bit numbering of Serpent according to its designers. We refer to (Anderson et al., 1998) for further details. This linear relation covers rounds 3 to 11 inclusive of the original 32-round Serpent; (iii) the linear relation involves only one E instance in Tandem-DM:

$$H^2_{i-1} \cdot \Gamma P \oplus E_{H^1_{i-1}\|m_i}(H^2_{i-1}) \cdot \Gamma C = (H^1_{i-1}\|m_i) \cdot \Gamma_1, \quad (14)$$

where Γ_1 is the bit mask for the key. Taking into account the feedforward of H^2_{i-1} and the fixed bit parity of $(H^1_{i-1}\|m_i) \cdot \Gamma_1$, (14) becomes $H^2_i \cdot \Gamma C = H^2_{i-1} \cdot (\Gamma P \oplus \Gamma C)$. Therefore, we can distinguish the compression function of Tandem-DM with 9-round Serpent instantiating E, using 2^{107} messages and equivalent effort.

Attack Abreast-DM Mode. We apply a linear attack on the compression function in Abreast-DM mode (8) with $n = 64$, $k = 128$ and E the PRESENT cipher. The attack proceeds as follows: we use 21-round PRESENT with 128-bit key in both instances of E. According to (Nakahara.Jr et al., 2009), the bias of the linear relation with mask $\Gamma =$

0000000000200000_x is $2^{-30.11}$, which leads to $2^{63.22}$ messages for a high success rate attack. We assume m_i and H^1_{i-1} are fixed as key input to the E instance, while H^2_{i-1} varies over $2^{63.22}$ values. We apply the linear approximation with bit mask Γ to both the input and the output of the compression function labeled by H^2_{i-1} and H^2_i. We obtain the linear relation:

$$H^2_{i-1} \cdot \Gamma \oplus E_{H^1_{i-1}\|m_i}(H^2_{i-1}) \cdot \Gamma = (H^1_{i-1}\|m_i) \cdot \Gamma_1, \quad (15)$$

where Γ_1 is the bit masks for the key $H^1_{i-1}\|m_i$ and H^1_0, H^2_0 are the initial values. We assume PRESENT with 128-bit keys, so there is no need for a transformation g for the key input in this case. The right-hand-side of (15) is a fixed parity bit since H^1_{i-1}, m_i and Γ_1 are fixed. Due to the feedforward of H^2_{i-1}, (15) can be simplified to $H^2_i \cdot \Gamma = 0$, that is, there is no more dependence on H^2_{i-1} nor on H^1_{i-1}. The distinguishing attack depends only on H^2_i. There is no linear relation involving the second E instance with H^1_i. In the original definition of Abreast-DM, the E instance whose input is H^1_{i-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. Therefore, using $2^{63.22}$ messages we can distinguish this compression function in Abreast-DM from an ideal mapping. Now, suppose that we instantiate E with 9-round Serpent with 256-bit key in Abreast-DM instead of Present. The attack would proceed very similarly to that described for Serpent, but with non iterative masks (ΓP, ΓC). The corresponding linear relation becomes $H^2_{i-1} \cdot \Gamma P \oplus E_{H^1_{i-1}\|m_i}(H^2_{i-1}) \cdot \Gamma C = (H^1_{i-1}\|m_i) \cdot \Gamma_1$, and due to the feedforward of H^2_{i-1}, it would simplify to $H^2_i \cdot \Gamma C = H^2_{i-1} \cdot (\Gamma P \oplus \Gamma C)$. In summary, we can distinguish the compression function of Abreast-DM with 9-round Serpent instantiating E using 2^{107} messages and equivalent effort.

Attack Parallel-DM Mode. The Parallel-DM is a double-block length hash mode designed by Hohl *et al.* in (W.Hohl et al., 1993). We apply a linear attack in this mode (6,7) on *the hash function* using 21-round PRESENT with $k = 80$ and $n = 64$. We assume there is a mapping $g : \mathbb{Z}_2^{64} \to \mathbb{Z}_2^{80}$ that transforms a 64-bit string to an 80-bit string. The precise description of g is not important. As long as the input of g is fixed, its output will be fixed as well, and vice-versa. In the attack we use messages of the form $M = m^1_1\|m^2_1\|m^1_2\|m^2_2$, where $m^1_2\|m^2_2$ contains (eventual) padding and the length of M. We assume m^1_2, m^2_2 and m^2_1 are fixed as key inputs to the E instances. But, we make m^1_1 assume all possible 2^{64} values. Since H^1_0 and H^2_0 are fixed, both $E_{g(m^1_1\oplus m^2_1)}(H^1_0 \oplus m^1_1)$ and $E_{g(m^1_1\oplus m^2_1)}(H^2_0 \oplus m^2_1)$ behave as random functions. According to (V.Rijmen et al., 1997), using all 2^{64} values of m^1_1 we expect to obtain about $2^{64}/e \approx 2^{62.56}$ distinct values from $E_{g(m^1_1\oplus m^2_1)}(H^1_0 \oplus m^1_1)$, where $e \approx 2.718$. According to (Nakahara.Jr et al., 2009), the bias of the linear relation with mask $\Gamma = 0000000000200000_x$ is $2^{-30.11}$, and this amount of plaintext allows to achieve a high success rate attack. Applying the bit mask Γ to both the input and the output of the compression function labeled by H^1_{i-1} and H^1_i we obtain the linear relation:

$$(H^1_1 \oplus m^1_2) \cdot \Gamma \oplus E_{g(m^1_2\oplus m^2_2)}(H^1_1 \oplus m^1_2) \cdot \Gamma = g(m^1_2 \oplus m^2_2) \cdot \Gamma_1, \quad (16)$$

where Γ_1 is the bit mask for the key $g(m^1_2 \oplus m^2_2)$. The right-hand-side of (16) is a fixed parity bit since m^1_2, m^2_2 and Γ_1 are fixed. Due to the feedforward of H^1_1, m^1_2 and same mask Γ for both input and output, (16) can be simplified to $H^1_2 \cdot \Gamma = 0$ that is, there is no more dependence on H^1_1 nor on m^1_2. Note that $H^1_2 = H^1_1 \oplus m^1_2 \oplus E_{g(m^1_2\oplus m^2_2)}(H^1_1 \oplus m^1_2)$. Thus, the distinguishing attack depends only on H^1_2 (half the hash digest). There is no linear relation involving the second E instance with H^2_i. Therefore, using 2^{64} messages we can distinguish this hash function in Parallel-DM from an ideal mapping.

Now, suppose we use 9-round Serpent with 128-bit key instantiating E. The attack would proceed similarly as in the previous paragraph, but (i) the attack is restricted to the compression function; (ii) the bias would be 2^{-52} and therefore, $8 \cdot (2^{-52})^{-2} = 2^{107}$ messages would be required for a high success rate attack. This means that $2^{107} \cdot e \approx 2^{108.44}$ values $H^1_{i-1}0$ will be needed; (iii) the bit masks are not iterative for the case of Serpent. We call the input and output masks simply ΓP and ΓC; the key mask is again denoted Γ_1. Their exact value can be found in (E.Biham et al., 2002); (iv) the linear relation involves only one E instance, and it would become $(H^1_{i-1} \oplus m^1_i) \cdot \Gamma P \oplus E_{m^1_i\|m^2_i}(H^1_{i-1} \oplus m^1_i) \cdot \Gamma C = (m^1_i\|m^2_i) \cdot \Gamma_1$. Taking into account the feedforward of H^1_{i-1} and the fixed bit parity of $(m^1_i\|m^2_i) \cdot \Gamma_1$, the linear relation becomes $(H^1_{i-1} \oplus m^1_i) \cdot (\Gamma P \oplus \Gamma C) = H^1_i \cdot \Gamma C$, which involves only inputs and outputs from the compression function. Therefore, we can distinguish the compression function of Parallel-DM with 9-round Serpent instantiating E using $2^{108.44}$ messages and equivalent effort.

5 CONCLUSIONS

This paper presented linear analyses of block-cipher based hash functions such as DM-PRESENT-80 and

Table 1: Attack complexities. Memory is negligible.

Target	Time	Mode
hash	2^{64}	DM-PRESENT-80 (1)
comp	$2^{49.84}$	DM-DES (2)
comp	$2^{63.22}$	MMO-PRESENT-80 (1)
comp	$2^{63.22}$	MP-PRESENT-80 (1)
comp	$2^{63.22}$	H-PRESENT-128 (3)
comp	$2^{63.22}$	Tandem-DM (3)
comp	2^{107}	Tandem-DM (4)
comp	$2^{63.22}$	Abreast-DM (3)
comp	2^{107}	Abreast-DM (4)
hash	2^{64}	Parallel-DM (1)
comp	$2^{108.44}$	Parallel-DM (5)

H-PRESENT-128. We demonstrated non-random properties of block ciphers also for compression functions in MMO and MP modes. Our attacks also included double-block-length hash modes such as Tandem-DM, Hirose's, Abreast-DM and Parallel-DM. Attack complexities are listed in Table 1. Notation: (1) 21-round PRESENT-80, (2) 16-round DES, (3) 21-round PRESENT-128, (4) 9-round Serpent-256, (5) 9-round Serpent-128. Based on these results we conclude that the DM and Parallel-DM modes are the weakest concerning linear attacks. These results also show that the Merkle-Damgård padding scheme used in DM mode is not enough to counter linear analysis, and thus avoid nonrandom detection attacks. It is well known that, if the Merkle-Damgård padding scheme is used, collision-resistance in the compression function propagates to the hash function (I.B.Damgård, 1989). On the other hand, our results show that, in the case of linear attacks aimed at the DM mode, the MD strengthening scheme was not effective to preclude nonrandom weaknesses to propagate from the underlying block cipher to the full hash function.

ACKNOWLEDGEMENTS

Research funded by INNOVIRIS, the Brussels Institute for Research and Innovation, under the ICT Impulse program CRYPTASC.

REFERENCES

A.Bogdanov, Knudsen, L., Leander, G., Paar, C., Poschmann, A., Robshaw, M., Seurin, Y., and Vikkelsoe, C. (2007). Present: an ultra-lightweight block cipher. In *9th Int. Workshop on Cryptographic Hardware and Enbedded Sysytems (CHES)*, LNCS 4727, pages 450–466. Springer.

A.Bogdanov, Leander, G., Paar, C., Poschmann, A., Robshaw, M., and Seurin, Y. (2008). Hash functions and rfid tags: mind the gap. In *CHES*, LNCS 5154, pages 283–299. Springer.

Anderson, R., Biham, E., and Knudsen, L. (1998). Serpent: a proposal for the advanced encryption standard. NIST AES proposal.

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

E.Biham, Dunkelman, O., and Keller, N. (2002). Linear cryptanalysis of reduced round serpent. In *Fast Software Encryption (FSE)*, LNCS 2355, pages 219–238. Springer.

FIPS (1993). Data encryption standard. Federal Info. Proc. Standards Pub. 46-2, supersedes FIPS PUB 46-1.

I.B.Damgård (1989). A design principle for hash functions. In *Adv. in Cryptology, Crypto'89*, LNCS 435, pages 416–427. Springer.

Lai, X. and Massey, J. (1993). Hash function based on block ciphers. In *Adv. in Cryptology, Eurocrypt'92*, LNCS 658, pages 55–70. Springer.

Matsui, M. (1994). The first experimental cryptanalysis of the data encryption standard. In *Adv. in Cryptology, Crypto 1994*, LNCS 839, pages 1–11. Springer.

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

Merkle, R. (1989). One way hash functions and des. In *Adv. in Cryptology, Crypto'89*, LNCS 435, pages 428–446. Springer.

M.Matsui (1994). Linear cryptanalysis method for des cipher. In *Adv. in Cryptology, Eurocrypt'93*, LNCS 765, pages 386–397. Springer.

Nakahara.Jr, J., Sepehrdad, P., Zhang, B., and Wang, M. (2009). Linear (hull) and algebraic cryptanalysis of the block cipher present. In *Cryptology and Network Security, CANS 2009*, LNCS 5888, pages 58–75. Springer.

NIST (2007). Announcing request for candidate algorithm nominations for a new cryptographic hash algorithm (sha-3) family. Federal Register, vol.72, no.212, Nov.2.

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

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

W.Hohl, Lai, X., Meier, W., and Waldvogel, C. (1993). Security of iterated hash functions based on block ciphers. In *Adv. in Cryptology, Crypto'93*, LNCS 773, pages 379–390. Springer.