Information Processing Letters 113 (2013) 301–306

Contents lists available at SciVerse ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Preimage and pseudo-collision attacks on step-reduced SM3 hash function ☆

Gaoli Wang *, Yanzhao Shen

School of Computer Science and Technology, Donghua University, Shanghai 201620, China

ARTICLE INFO

Article history:
Received 12 November 2012
Received in revised form 26 January 2013
Accepted 6 February 2013
Available online 8 February 2013
Communicated by D. Pointcheval

Keywords:
Cryptography
Preimage attack
Collision attack
Differential meet-in-the-middle
SM3
Hash function

ABSTRACT

SM3 [12] is the Chinese cryptographic hash standard which was announced in 2010 and designed by Wang et al. It is based on the Merkle–Damgård design and its compression function can be seen as a block cipher used in Davies–Meyer mode. It uses message block of length 512 bits and outputs hash value of length 256 bits.

This letter studies the security of SM3 hash function against preimage attack and pseudo-collision attack by using the weakness of diffusion process and linear message expansion. We propose preimage attacks on 29-step and 30-step SM3, and pseudo-preimage attacks on 31-step and 32-step SM3 out of 64 steps. The complexities of these attacks are 2^{245} 29-step operations, $2^{251.1}$ 30-step operations, 2^{245} 31-step operations and $2^{251.1}$ 32-step operations, respectively. These (pseudo-)preimage attacks are all from the 1-st step of the reduced SM3. Furthermore, these (pseudo-)preimage attacks can be converted into pseudo-collision attacks on SM3 reduced to 29 steps, 30 steps, 31 steps and 32 steps with complexities of 2^{122}, $2^{125.1}$, 2^{122} and $2^{125.1}$ respectively. As far as we know, the previously best known preimage attacks on SM3 cover 28 steps (from the 1-st step) and 30 steps (from the 7-th step).

1. Introduction

Cryptographic hash functions play an important role in modern cryptography. From the cryptographic viewpoint, a cryptographic hash function has to fulfill several security properties such as classical ones: collision resistance, preimage resistance and second preimage resistance. With the breakthroughs in the collision attacks on a series of standard hash functions such as MD5, SHA-0 and SHA-1 [2,13,14], preimage attack has drawn a great amount of attention from many researchers (see [1,4,6,7,9,10,15] for example). Up to now, the meet-in-the-middle technique [1,3] and many improved techniques such as initial structure, splice-and-cut, biclique and so on have been widely used in the preimage attack. Recently, a differential view on the meet-in-the-middle technique [6] was proved very useful for the preimage attack on hash functions with linear message expansion and weak diffusion properties.

SM3 [12] hash function is the Chinese cryptographic hash standard which was designed by Wang et al. and announced in 2010. The structure of SM3 resembles the structure of SHA-256. However, it has a more complex step function and stronger message dependency than SHA-256. Few attacks were published on SM3 hash function. The work in [16] presented preimage attacks on 28-step (from the 1-st step) and 30-step (from the 7-th step) SM3 with complexities of 2^{249} and $2^{241.5}$ respectively. In [16] it is stated that attacking steps from the 1-st step is difficult because the absorption property of the boolean functions within the first 16 steps does not hold. According to the linear message expansion and weakness in diffusion of SM3, we can find proper linear spaces and construct differential characteristics with high probabilities,

☆ This work was supported by the National Natural Science Foundation of China under Grant No. 61103238, the Fundamental Research Funds for the Central Universities.

* Corresponding author.
E-mail addresses: wanggaoli@dhu.edu.cn (G. Wang), yanzhao_shen@yahoo.com.cn (Y. Shen).

0020-0190/$ – see front matter
http://dx.doi.org/10.1016/j.ipl.2013.02.006

Table 1
Summary of the attacks on SM3 compression function (CF) and hash function (HF).

Attack	CF/HF	Steps	Time	Source
P. A.	HF	28	$2^{241.5}$	[16]
P. A.	HF	30[a]	2^{249}	[16]
B. A.	CF	32	$2^{14.4}$	[5]
B. A.	CF	33	$2^{32.4}$	[5]
B. A.	CF	34	$2^{53.1}$	[5]
B. A.	CF	35	$2^{117.1}$	[5]
C. A.	HF	20	Practical	[17]
P. C. A.	HF	24	Practical	[17]
P. A.	HF	29	2^{245}	Section 4
P. A.	HF	30	$2^{251.1}$	Section 4
P. P. A.	HF	31	2^{245}	Section 4
P. P. A.	HF	32	$2^{251.1}$	Section 4
P. C. A.	HF	29	2^{122}	Section 4
P. C. A.	HF	30	$2^{125.1}$	Section 4
P. C. A.	HF	31	2^{122}	Section 4
P. C. A.	HF	32	$2^{125.1}$	Section 4

[a] The attack starts from the 7-th step.
B. A. – Boomerang attack.
P. A. – Preimage attack.
C. A. – Collision attack.
P. P. A – Pseudo-preimage attack.
P. C. A. – Pseudo- or free-start collision attack.

then apply the differential meet-in-the-middle preimage attacks on step-reduced SM3 hash function. The probabilities of the differential characteristics are not influenced by the properties of the boolean functions. Therefore, the differential meet-in-the-middle preimage attacks on step-reduced SM3 we proposed can be from the 1-st step. Recently, a boomerang attack on SM3 reduced to 35 steps [5] was published with a complexity of $2^{117.1}$. A collision attack on 20-step SM3 and a pseudo-collision attack on 24-step SM3 [17] were proposed with practical complexity.

In this letter, we focus on the security evaluation of the preimage resistance and collision resistance of SM3 hash function. By applying the differential meet-in-the-middle technique and biclique technique to the analysis of SM3, we successfully present (pseudo-)preimage attacks and pseudo-collision attacks on 29-step, 30-step, 31-step and 32-step SM3 hash function. All of these attacks start from the 1-st step of SM3. This result provides a better understanding concerning the message expansion and diffusion properties of SM3 hash function. The previous results and the summary of our results are given in Table 1.

The rest of this letter is organized as follows. Section 2 introduces the techniques used throughout the letter. Section 3 gives a brief description of SM3 and some notations used in this letter. Section 4 presents preimage and pseudo-collision attacks on step-reduced SM3. Section 5 concludes this letter.

2. Techniques for preimage attack and pseudo-collision attack

In this section, we will introduce the related techniques used throughout the letter.

2.1. The meet-in-the-middle preimage attack

The meet-in-the-middle preimage attack [3] is a type of birthday attack and makes use of a space–time tradeoff. Firstly, the function is divided into two subparts E_f (forward process) and E_b (backward process). Then two set of values at the matching point are calculated by E_f and E_b respectively. The two computation procedures must be independent on each other so that the birthday attack rule can be applied. Lastly, check if there exists a value in one set that matches a value in the other set. The meet-in-the-middle technique can be combined with many techniques such as initial structure technique, splice-and-cut technique, biclique technique, etc. to improve the preimage attack.

2.2. The differential meet-in-the-middle technique

We review the differential meet-in-the-middle preimage attack [6] which uses the truncated differential [8] in the following. For a compression function $CF = E(M, IV) \oplus IV$, where $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ is a block cipher with block length n and key length k ($k > n$), let $T \in \{0,1\}^n$ be a truncation mask vector, and the equation $A =_T B$ denotes $T \wedge (A \oplus B) = 0$. Furthermore, the hash value H is given. Then the differential meet-in-the-middle preimage attack can be constructed as follows.

- Split the block cipher E into two parts, i.e., $E = E_2 \cdot E_1$, and find the linear spaces D_1, D_2 which satisfy $D_1 \cap D_2 = \{0\}$.
- For each $\delta_1 \in D_1$, search for the difference Δ_1, such that the equation $\Delta_1 =_T E_1(M, IV) \oplus E_1(M \oplus \delta_1, IV)$ holds with high probability for uniformly chosen message M.
- For each $\delta_2 \in D_2$, search for the difference Δ_2, such that the equation $\Delta_2 =_T E_2^{-1}(M, H \oplus IV) \oplus E_2^{-1}(M \oplus \delta_2, H \oplus IV)$ holds with high probability for uniformly chosen message M.
- Search for the candidate preimage using Algorithm 1.

Assume $p_1 = Pr[\Delta_1 =_T E_1(M, IV) \oplus E_1(M \oplus \delta_1, IV)]$ and $p_2 = Pr[\Delta_2 =_T E_2^{-1}(M, H \oplus IV) \oplus E_2^{-1}(M \oplus \delta_2, H \oplus IV)]$, then a true preimage can be obtained with probability $p_1 \cdot p_2$. If D_1 and D_2 both have dimension d, for a random M, the set $M \oplus D_1 \oplus D_2$ contains $2^{2d} = 2^d \times 2^d$ different messages. Using Algorithm 1, we can observe that a preimage can be obtained with a complexity of $(2^{n-d}\Gamma + 2^{n-t}\Gamma_{re})/(p_1 \cdot p_2)$, where Γ is the cost of one compression function operation, Γ_{re} is the cost of retesting a candidate preimage and t is the hamming weight of T. For the detailed description, we refer to [6].

2.3. Converting pseudo-preimage attack into pseudo-collision attack

The work in [11] proposed a technique to convert preimage attack into pseudo-collision attack. Assume we can get a t-bit partial target preimage M with matching point in the last step with complexity 2^k, then by finding $2^{(n-t)/2}$ different t-bit partial target preimages Ms, we can get a pseudo-collision with the complexity of $2^{(n-t)/2} \times 2^k$. If the t-bit partial target preimages are constructed by the meet-in-the-middle technique, then the complexity of the pseudo-collision attack can be improved. For example, in the case of $t = 6$ and $|m_1| = |m_2| = 5$ ($> t/2$), where $|m_1|$