Practical Collision Attacks on 22- and 23-Step SM3 (Preliminary Version)

Sitong Yuan, Haorui Cui and Hongbo Yu*

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
yuanst23@mails.tsinghua.edu.cn,cuihr26@mails.tsinghua.edu.cn
yuhongbo@mail.tsinghua.edu.cn

Abstract. SM3 is a 256-bit cryptographic hash function specified in the Chinese national standard GB/T 32905-2016 and the international standard ISO/IEC 10118-3:2018. In this paper, we present practical collisions for 22- and 23-step SM3 with the standard initial value. Compared with the 20-step collision result of Mendel et al. (2013), our constructions cover two and three additional steps, respectively. The time complexities are approximately $2^{52.4}$ and $2^{54.5}$ evaluations of 22- and 23-step SM3, respectively. We also present a semi-free-start collision for 24-step SM3.

Keywords: SM3 · collisions · semi-free-start collisions

1 Introduction

SM3 was designed by Xiaoyun Wang and her collaborators and released by the State Cryptography Administration of China in 2010 [Sta10]. It was subsequently standardized as GB/T 32905-2016 in China [Sta16] and incorporated into ISO/IEC 10118-3:2018 [ISO18]. An English specification is available in the official translation of GM/T 0004-2012 [Cry12]. SM3 is widely used in commercial cryptographic products deployed in finance, electric power, transportation, and e-government.

SM3 combines strong diffusion with a complex step function, and each step incorporates two message-word inputs, W_j and the derived word W'_j. These features pose challenges for constructing differential characteristics and conforming message pairs.

In 2013, Mendel et al. presented practical collisions for 20-step SM3 with the standard initial value [MNS13]. The attack uses a common first message block to connect the standard initial value to the collision construction in the second block. Table 1 summarizes previous collision attacks on SM3 and our results.

Table 1: Collision attacks on reduced SM3

Attack type	Steps	Time	Space	Reference
Collision	20	practical	—	[MNS13]
Collision	22	$2^{52.4}$ (practical)	—	Section 4
Collision	23	$2^{54.5}$ (practical)	—	Section 5
Free-start collision	24	practical	—	[MNS13]
Semi-free-start collision	24	practical	—	Section 6

*Corresponding author: yuhongbo@mail.tsinghua.edu.cn.

2 Description of SM3

2.1 Message Expansion

The SM3 compression function consists of a message expansion and a state update. A 512-bit message block is divided into sixteen 32-bit words m_0, ..., m_{15}. We use + for addition modulo 2^{32} and $\lll$ for cyclic left rotation. Steps are indexed starting from zero. The message expansion starts with

$$W_j = m_j, \qquad 0 \le j < 16. \tag{1}$$

For $16 \le j < 68$, the expanded message words satisfy the recurrence

$$W_j = P_1(W_{j-16} \oplus W_{j-9} \oplus (W_{j-3} \lll 15)) \oplus (W_{j-13} \lll 7) \oplus W_{j-6}. \tag{2}$$

Each step uses W_j and a second derived message word W'_j, defined by

$$W'_j = W_j \oplus W_{j+4}, \qquad 0 \le j < 64. \tag{3}$$

The two linear permutations are defined as

$$\begin{aligned} P_0(X) &= X \oplus (X \lll 9) \oplus (X \lll 17), \\ P_1(X) &= X \oplus (X \lll 15) \oplus (X \lll 23). \end{aligned} \tag{4}$$

2.2 State Update

Let A, B, C, D, E, F, G, and H be 32-bit word registers, and $SS1$, $SS2$, $TT1$, and $TT2$ be intermediate variables. For each message block, initialize the registers with the input chaining value $V^{(i)}$ [Cry12]:

$$A \| B \| C \| D \| E \| F \| G \| H \leftarrow V^{(i)}. \tag{5}$$

For $j = 0, \ldots, 63$, perform the following operations in order. All additions are modulo 2^{32}:

$$\begin{aligned} SS1 &\leftarrow \big((A \lll 12) + E + (T_j \lll (j \bmod 32))\big) \lll 7, \\ SS2 &\leftarrow SS1 \oplus (A \lll 12), \\ TT1 &\leftarrow \mathrm{FF}_j(A, B, C) + D + SS2 + W'_j, \\ TT2 &\leftarrow \mathrm{GG}_j(E, F, G) + H + SS1 + W_j, \\ D &\leftarrow C, \\ C &\leftarrow B \lll 9, \\ B &\leftarrow A, \\ A &\leftarrow TT1, \\ H &\leftarrow G, \\ G &\leftarrow F \lll 19, \\ F &\leftarrow E, \\ E &\leftarrow P_0(TT2). \end{aligned} \tag{6}$$

The Boolean functions and step constants are defined as

$$\mathrm{FF}_j(x,y,z) = \begin{cases} x \oplus y \oplus z, & 0 \le j < 16, \\ (x \wedge y) \vee (x \wedge z) \vee (y \wedge z), & 16 \le j < 64, \end{cases}$$

$$\mathrm{GG}_j(x,y,z) = \begin{cases} x \oplus y \oplus z, & 0 \le j < 16, \\ (x \wedge y) \vee (\neg x \wedge z), & 16 \le j < 64, \end{cases} \tag{7}$$

$$T_j = \begin{cases} \texttt{79cc4519}, & 0 \le j < 16, \\ \texttt{7a879d8a}, & 16 \le j < 64. \end{cases}$$

After all 64 steps, the output chaining value is computed as

$$V^{(i+1)} = (A\|B\|C\|D\|E\|F\|G\|H) \oplus V^{(i)}. \tag{8}$$

Let $C_r(h, M)$ denote the compression function retaining the first r steps and including feedforward. Throughout this paper, one SM3 evaluation means one evaluation of the corresponding reduced compression function.

3 Message-Difference Structures

We restrict nonzero differences to a small number of message words and require cancellation in the subsequent expanded words. This allows a local collision to be constructed in the first few steps, followed by zero differences in all remaining steps. Table 2 shows the message words used for 22, 23, and 24 steps and their dependencies in the message expansion.

The 22- and 23-step structures apply to the second block of a two-block construction with a common first block. The 24-step structure applies to a single compression-function evaluation with a common, freely chosen initial value.

Table 2: Message-difference structures for 22- and 23-step collisions and a 24-step SFS collision

22 steps (collision)

$j\backslash i$	0	1	3	4	7	10
0	•					
1		•				
2						
3			•			
4				•		
5						
6						
7					•	
8						
9						
10						•
11						
12						
13						
14						
15						
16	•		•		•	•
17		•		•		
18						
19			•			•
20				•	•	
21						
22						
23					•	•
24						
25						

23 steps (collision)

$j\backslash i$	1	2	4	5	8	11
0						
1	•					
2		•				
3						
4			•			
5				•		
6						
7						
8					•	
9						
10						
11						•
12						
13						
14						
15						
16						
17	•		•		•	•
18		•		•		
19						
20			•			•
21				•	•	
22						
23						
24					•	•
25						
26						

24 steps (SFS)

$j\backslash i$	0	2	3	5	6	9	12
0	•						
1							
2		•					
3			•				
4							
5				•			
6					•		
7							
8							
9						•	
10							
11							
12							•
13							
14							
15							
16	•		•				
17							
18		•		•		•	•
19			•		•		
20							
21				•			•
22					•	•	
23							
24							
25						•	•
26							
27							

Column i corresponds to a message word W_i with nonzero difference. Shaded rows have nonzero message differences. A • marks an occurrence of the column word in the expansion equation for W_j. For each unshaded row marked with bullets, the message-word differences must cancel.

4 Collisions for 22-Step SM3

For 22-step SM3, we use a two-block collision construction. Let h_0 be the standard initial value. We first select a common message block M_0 and compute the input chaining value h_1 for the second block. We then construct M_1 and M_1' according to the difference structure in Section 3. The resulting message pair satisfies

$$\begin{aligned} h_1 &= C_{22}(h_0, M_0), \\ C_{22}(h_1, M_1) &= C_{22}(h_1, M_1') = h_2, \qquad M_1 \neq M_1'. \end{aligned} \tag{9}$$

Table 3 gives the colliding message pair and the intermediate chaining value. Each 32-bit word is represented by eight hexadecimal digits, and the message words are in big-endian order. The value h_2 is the output after compressing the second block, including feedforward.

Table 3: A collision for 22-step SM3

h_0	7380166f 4914b2b9 172442d7 da8a0600 a96f30bc 163138aa e38dee4d b0fb0e4e
M_0	a96c1688 0cf030f5 60c52a29 a03e9f02 2804250c f71ac774 24bb8b84 279e110b dab4cf9a 308e4398 413c2caf 0ae6ea73 b19a981c e5886107 60d39901 cccce42b
h_1	c049858f c9065f9c 307aaac8 8284aeca 6e9ff9e1 cd668e62 b5c64db7 291f683c
M_1	6b648c68 84476675 e033ccc8 e6caecb3 d9538ca9 c1e4eb99 7fff1787 55374582 91079596 c540139a 1283f8ce 11af479e 205c2abd e326d25d 4231e15e 5c1f901d
M_1'	6b6c886c 24a74715 e033ccc8 e6c0ecb1 d9538cab c1e4eb99 7fff1787 51354782 91079596 c540139a 1289f8cc 11af479e 205c2abd e326d25d 4231e15e 5c1f901d
$h_2 = h_2'$	a26aee58 365c4672 8d23941d a3717024 562fdafe 0055237a 944e020d 4ed48936

The time complexity of finding this collision is approximately $2^{52.4}$ evaluations of 22-step SM3.

5 Collisions for 23-Step SM3

For 23-step SM3, we likewise start from the standard initial value h_0, use a common block M_0 to produce h_1, and construct the second blocks M_1 and M_1' according to the structure in Section 3. The two computations have identical states at the end of step 11 and zero differences throughout the remaining steps, giving

$$\begin{aligned} h_1 &= C_{23}(h_0, M_0), \\ C_{23}(h_1, M_1) &= C_{23}(h_1, M_1') = h_2, \qquad M_1 \neq M_1'. \end{aligned} \tag{10}$$

The colliding message pair is given in Table 4, together with the standard initial value h_0, the intermediate chaining value h_1, and the compression output h_2.

The time complexity of finding this collision is approximately $2^{54.5}$ evaluations of 23-step SM3.

Table 4: A collision for 23-step SM3

h_0	7380166f 4914b2b9 172442d7 da8a0600 a96f30bc 163138aa e38dee4d b0fb0e4e
M_0	d22c8eaa c24b03cb 24ab41ee 16ad81fe 7cfa7132 59a23ec7 83b181f6 ffa99d5f c462d074 45255fb8 abe00fff 9467cef6 97468c92 82f3ead1 51dfcf08 117d41fa
h_1	a808fb4c a3ac6317 4753e6c0 7f6fcfaf b9debb47 d3ad7306 08f1688f d9c2335d
M_1	10642197 2a406151 ea4b13c2 816d75ee b33d73a6 eeaf5e1a b0b5d9c0 53391648 db9d6407 07ab3e7f 97780cca 49df85de ef9d7e78 a2450cd0 1b424dbc 5f5f855e
M_1'	10642197 2b406350 b2632bca 816d75ee b3bd7126 ee2f5e1a b0b5d9c0 53391648 5b9c6487 07ab3e7f 97780cca 495f875e ef9d7e78 a2450cd0 1b424dbc 5f5f855e
$h_2 = h_2'$	07edcc34 7c75d378 891e8f39 bb671548 193513a7 ed1181de aa0e4108 e3e11f7c

6 A Semi-Free-Start Collision for 24-Step SM3

For a semi-free-start collision, the two distinct message blocks use the same freely chosen initial value h. Unlike the collisions in Sections 4 and 5, this construction does not require h to be the standard initial value. We obtain message blocks M and M' satisfying

$$C_{24}(h, M) = C_{24}(h, M') = h_{\text{out}}, \qquad M \neq M'. \tag{11}$$

The state differences vanish at the end of step 12. The message words W_j and W_j' have zero differences for $13 \leq j \leq 23$, so the two computations remain identical through step 23. Since the initial values are equal, the feedforward outputs are also equal.

Table 5 gives the common initial value, the two message blocks, and their common compression output after feedforward. All values are written as 32-bit hexadecimal words in big-endian order.

We obtained this semi-free-start collision through approximately 20 minutes of SAT search.

Table 5: A semi-free-start collision for 24-step SM3

h	8f0b515b 7d683d59 d9ad89a7 21be1c46 40139fa1 f8cd729b 1a49249f 9fa2566b
M	cf839763 e498bdd8 d0fda5e3 d61fdebb ae941f30 7c97b227 91376304 8004e999 737a6778 2649bcff 2923c186 34699b5c 5281d54a 96764093 0f188746 ccd7929a
M'	c783bf63 e498bdd8 d0ffa4e2 fe27d6e3 ae941f30 fc953227 11376304 8004e999 737a6778 27493c7f 2923c186 34699b5c d283554a 96764093 0f188746 ccd7929a
h_{out}	3f0c60cf 278c38b4 b4e0c7bf 163ccd52 4e0f56e5 f91b12b5 60d90a47 122a7264

References

[Cry12] Cryptography Standardization Technical Committee of China. SM3 Cryptographic Hash Algorithm. GM/T 0004-2012, 2012. Official English translation made available in November 2024. http://www.gmbz.org.cn/upload/2024-11-18/1731899426565012428.pdf.

[ISO18] ISO/IEC. IT Security techniques—Hash-functions—Part 3: Dedicated hash-functions. ISO/IEC 10118-3:2018, 2018. https://www.iso.org/standard/67116.html.

[MNS13] Florian Mendel, Tomislav Nad, and Martin Schläffer. Finding Collisions for Round-Reduced SM3. In *Topics in Cryptology – CT-RSA 2013*, volume 7779 of *Lecture Notes in Computer Science*, pages 174–188. Springer, 2013. https://doi.org/10.1007/978-3-642-36095-4_12.

[Sta10] State Cryptography Administration of China. Announcement on the release of the SM3 cryptographic hash algorithm. Announcement No. 22, December 17, 2010 (in Chinese), 2010. https://sca.gov.cn/sca/xwdt/2010-12/17/content_1002389.shtml.

[Sta16] Standardization Administration of China. Information security techniques—SM3 cryptographic hash algorithm. GB/T 32905-2016, 2016. https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=45B1A67F20F3BF339211C391E9278F5E.