Zhao JY, Wang MQ, Wen L. Improved linear cryptanalysis of CAST-256. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 29(6): 1134–1139 Nov. 2014. DOI 10.1007/s11390-014-1496-8

Improved Linear Cryptanalysis of CAST-256

Jing-Yuan Zhao (赵静远), Mei-Qin Wang* (王美琴), and Long Wen (温　隆)

Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, Shandong University Jinan 250100, China

School of Mathematics, Shandong University, Jinan 250100, China

E-mail: jingyuanzhao@mail.sdu.edu.cn; mqwang@sdu.edu.cn; longwen@mail.sdu.edu.cn

Received July 16, 2013; revised April 1, 2014.

Abstract CAST-256, a first-round AES (Advanced Encryption Standard) candidate, is designed based on CAST-128. It is a 48-round Generalized-Feistel-Network cipher with 128-bit block accepting 128, 160, 192, 224 or 256 bits keys. Its S-boxes are non-surjective with 8-bit input and 32-bit output. Wang *et al.* identified a 21-round linear approximation and gave a key recovery attack on 24-round CAST-256. In ASIACRYPT 2012, Bogdanov *et al.* presented the multidimensional zero-correlation linear cryptanalysis of 28 rounds of CAST-256. By observing the property of the concatenation of forward quad-round and reverse quad-round and choosing the proper active round function, we construct a linear approximation of 26-round CAST-256 and recover partial key information on 32 rounds of CAST-256. Our result is the best attack according to the number of rounds for CAST-256 without weak-key assumption so far.

Keywords CAST-256, linear cryptanalysis, block cipher, Generalized-Feistel-Network

1 Introduction

Differential cryptanalysis[1] and linear cryptanalysis[2] are two basic methods for evaluating the security of block ciphers. New designed block ciphers consider resistance to them. For example, CAST-256 uses non-bijective S-boxes which transform small input to large output to resist to the differential cryptanalysis, and produces S-boxes with bent function to resist to linear cryptanalysis. Many new variants of them have been developed such as truncated differential cryptanalysis[3], impossible differential cryptanalysis[4-5], multiple differential cryptanalysis[6-7], boomerang attack[8], differential-algebraic cryptanalysis[9-10], multiple linear cryptanalysis[11-13] and zero-correlation linear cryptanalysis[14]. However, the differential and the linear cryptanalysis are still very important because the best attacks for some block ciphers use the differential or linear cryptanalysis.

CAST-256[15] is a candidate of AES (Advanced Encryption Standard)① which is an extension of ISO block cipher CAST-128[16]. It is a Generalized-Feistel-Network block cipher with three different round functions, F_1, F_2 and F_3, and consists of six forward quad-rounds and six reverse quad-rounds. The total round number is 48. The block size is 128 bits and the key size can be 128, 160, 192, 224 or 256 bits.

CAST-256 has been actively attacked by the cryptanalysts. Nakahara and Rasmussen identified some 12-round linear approximations and gave a distinguishing attack on 12-round CAST-256[17]. Wang *et al.*[18] constructed a 21-round linear approximation to recover the key of 24-round CAST-256. Sun *et al.* proposed an optimized searching algorithm for the linear approximation of round function and found a new linear approximation for the round function F_2 with a slightly better bias[19] than that identified by Wang *et al.*[18], which cannot be used to attack more rounds. Wagner proposed the boomerang attack on 16-round CAST-256[8]. Biham claimed that there were 20-round impossible differentials[20]. Seki and Kaneko gave a differential attack on 36-round CAST-256 under a weak-key assumption that covers 2^{-35} of the keys[21]. Bogdanov *et al.* attacked 28 rounds of CAST-256 using

---

Regular Paper

This work was supported by the National Basic Research 973 Program of China under Grant No. 2013CB834205, the National Natural Science Foundation of China under Grant Nos. 61133013, 61070244 and 61103237, the Program for New Century Excellent Talents in University of China under Grant No. NCET-13-0350, as well as the Interdisciplinary Research Foundation of Shandong University under Grant No. 2012JC018.

*Corresponding Author

① Specification for the advanced encryption standard, November 2001. http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf, Sept. 2014.

multidimensional zero-correlation linear cryptanalysis in ASIACRYPT 2012[22].

Linear cryptanalysis is typically a known-plaintext or a ciphertext-only attack proposed by Matsui[2] and uses a linear approximation as a distinguisher. The linear approximation consists of a linear combination of plaintext, ciphertext and key bits, holding with a relatively high parity deviation from the uniform parity distribution. The effectiveness of a linear approximation is evaluated by a parameter called bias, denoted as ϵ, which is the absolute value of difference between the probability of a linear approximation and $\frac{1}{2}$. For the linear cryptanalysis proposed by Matsui, the higher the bias is, the more attractive the linear approximations are, since they require less plaintext-ciphertext pairs. However, zero-correlation linear cryptanalysis proposed by Bogdanov and Rijmen[14] uses linear approximations with zero bias. These linear approximations allow us to distinguish a cipher from a random permutation, or to recover subkey bits.

In this paper, we find that the concatenation of forward quad-round and reverse quad-round can be used to produce longer linear approximations, and thus we put the concatenation of forward quad-round and reverse quad-round inside the linear distinguisher. Moreover, the linear approximation of CAST-256 can be iteratively produced from the linear approximation of the round function F_1, F_2 or F_3, thereby we will decide which round function's linear approximation can be used to produce a better linear distinguisher for CAST-256. From [18], the linear approximation of a quad-round with F_2 is better than that with F_1 or F_3. However, we found that if the linear approximations of F_2 and F_3 are used, the linear distinguisher of 24-round and 22-round CAST-256 can be constructed, respectively. If the linear approximation of F_1 is used, the linear distinguisher of 26-round CAST-256 can be discovered, with which we can recover partial key information for 32 rounds of CAST-256. Although we cannot recover the whole key, the partial key information recovery attack is still significant to some extent in cryptography. In this way, our attack is the best known attack on CAST-256 according to the number of rounds without the weak-key assumption. Table 1 is the summary and comparison of attacks on CAST-256.

This paper is organized as follows. Section 2 describes the algorithm of CAST-256. The linear approximations of CAST-256 are derived in Section 3. Section 4 gives the partial key recovery attack on 32-round CAST-256. We conclude this paper in Section 5.

2 CAST-256 Algorithm

CAST-256[15], a first-round AES candidate②, is a block cipher published in June 1998. It is an extension of CAST-128[16]. Both of them were designed according to the "CAST" design methodology invented by Adams[23], which uses three kinds of round functions based on 8×32 S-boxes. The block size of CAST-256 is 128 bits and the key size can be 128, 160, 192, 224 or 256 bits. The number of rounds is 48 for any key size. The design is based on a Generalized-Feistel-Network with four branches and consists of six forward quad-rounds followed with six reverse quad-rounds.

Denote the three kinds of round function as F_1, F_2 and F_3. $I = \{I_1|I_2|I_3|I_4\}$ is the 32-bit input of the round function, S_i, $1 \leqslant i \leqslant 4$, is the i-th S-box of the round function, and O is the 32-bit output of the round function. We use "+" and "−" to denote the addition and subtraction modulo 2^{32}, respectively, "$\oplus$" is bitwise exclusive-OR and "$\lll$" means the left rotation. We can describe F_1, F_2 and F_3 as follows:

$$\begin{aligned}
F_1 : I &= ((k_m + I) \lll k_r),\\
O &= ((S_1[I_1] \oplus S_2[I_2] - S_3[I_3]) + S_4[I_4]);\\
F_2 : I &= ((k_m \oplus I) \lll k_r),\\
O &= ((S_1[I_1] - S_2[I_2] + S_3[I_3]) \oplus S_4[I_4]);\\
F_3 : I &= ((k_m - I) \lll k_r),\\
O &= ((S_1[I_1] + S_2[I_2] \oplus S_3[I_3]) - S_4[I_4]),
\end{aligned}$$

where k_r and k_m are the 5-bit "rotation" subkey and the 32-bit "masking" subkey for current round, respectively.

Table 1. Summary of Attacks on CAST-256

Attack	Number of Rounds	Key Size	Data	Time	Memory (Byte)	Ratio of Weak Keys
Distinguishing[17]	12	128	$2^{101.0}$KP	$2^{101.00}$	2^{103}	1
Boomerang[8]	16	128	$2^{49.3}$CP	-	-	1
Differential[21]	36	256	$2^{123.0}$CP	$2^{182.00}$	-	2^{-35}
Linear[18]	24	192	$2^{124.1}$KP	$2^{156.52}$	-	1
Multidim. ZC[22]	28	256	$2^{98.8}$KP	$2^{246.90}$	2^{68}	1
Linear (Ours)	32	256	$2^{126.8}$KP	$2^{251.00}$	2^{99}	1

②Specification for the advanced encryption standard, November 2001. http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf, Sept. 2014.

Based on the defined F_i $(1 \leqslant i \leqslant 3)$, let $\beta = (A, B, C, D)$ be a 128-bit block where A, B, C, D are 32-bit words for the inputs of different round functions F_i $(1 \leqslant i \leqslant 3)$. Define the "forward quad-round" as $\beta = Q(\beta)$:

$$C = C \oplus F_1(D, k_{r1}^i, k_{m1}^i),$$
$$B = B \oplus F_2(C, k_{r2}^i, k_{m2}^i),$$
$$A = A \oplus F_3(B, k_{r3}^i, k_{m3}^i),$$
$$D = D \oplus F_1(A, k_{r4}^i, k_{m4}^i),$$

and the "reverse quad-round" as $\beta = Q'(\beta)$:

$$D = D \oplus F_1(A, k_{r1}^i, k_{m1}^i),$$
$$A = A \oplus F_3(B, k_{r2}^i, k_{m2}^i),$$
$$B = B \oplus F_2(C, k_{r3}^i, k_{m3}^i),$$
$$C = C \oplus F_1(D, k_{r4}^i, k_{m4}^i),$$

where k_{rj}^i and k_{mj}^i $(1 \leqslant j \leqslant 4, 1 \leqslant i \leqslant 12)$ are the rotation subkey and the masking subkey in the j-th round of the i-th quad-round, respectively.

One forward quad-round and one reverse quad-round of CAST-256 are shown in Fig.1 and Fig.2, respectively. The concatenation of forward quad-round and reverse quad-round is shown in Fig.3.

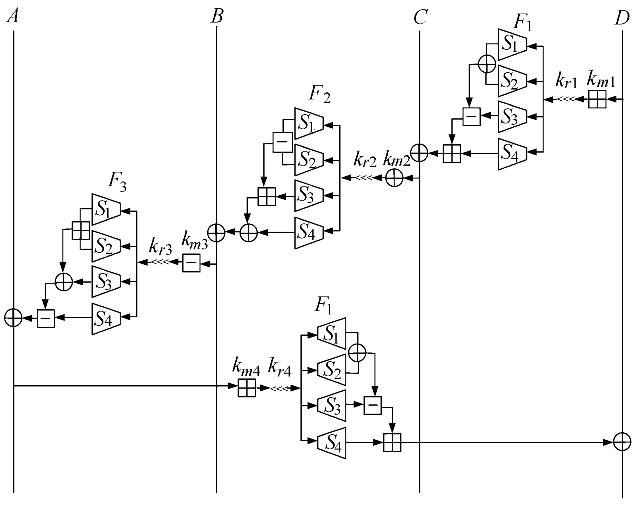

Fig.1. Forward quad-round of CAST-256.

3 Linear Approximation of CAST-256

As CAST-256 uses Generalized-Feistel-Network, the linear approximation of one-round function with zero input mask and non-zero output mask has more advantage to cover more rounds than that with both non-zero input and output masks. Fortunately, since the S-boxes of CAST-256 are non-surjective bent functions with 8-bit input and 32-bit output, the bias of the linear approximation $(0 \rightarrow \Gamma')$ (0 is the input mask and Γ' is

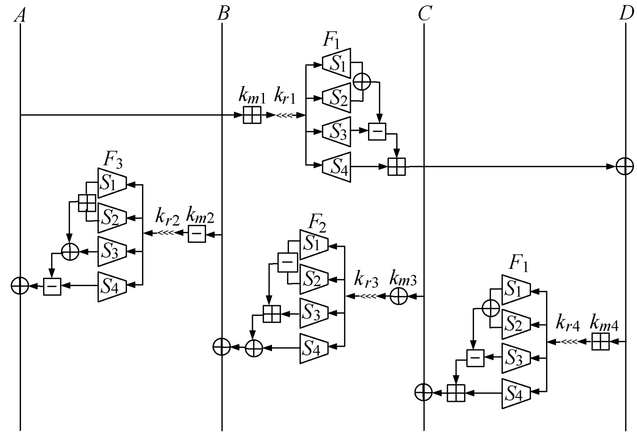

Fig.2. Reverse quad-round of CAST-256.

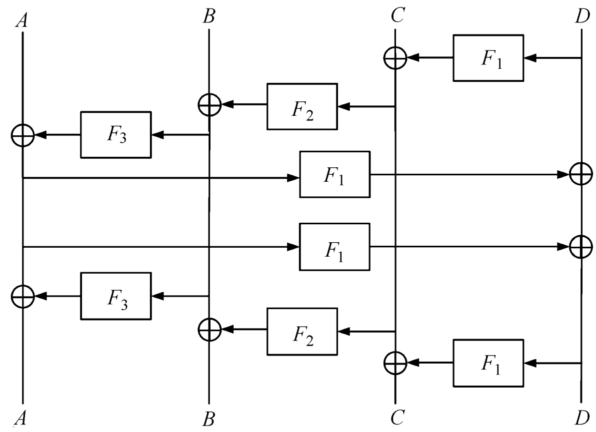

Fig.3. Concatenation of forward quad-round and reverse quad-round.

the output mask) of them is non-zero. Actually there are linear approximations $(0 \rightarrow \Gamma)$ with non-zero bias for the round functions F_1, F_2 and F_3.

Based on these observations, Wang *et al.* searched the linear approximations $(0 \rightarrow \Gamma)$ of the round functions F_1, F_2 and F_3, where the Hamming-weight of Γ is less than 6 for F_2 and the Hamming-weight of Γ is less than 4 for F_1 and F_3[18]. As a result, the best known linear approximation for the round function F_2 is $(0 \rightarrow 03400000_x)$ with a bias of $2^{-12.91}$, which had been used to produce the 21-round linear approximation and attack 24 rounds of CAST-256. Then Sun *et al.*[19] optimized the searching algorithm and searched all the linear approximations $(0 \rightarrow \Gamma)$ of the round functions F_1, F_2 and F_3 for all possible values of Γ. As a result, they only identified a better linear approximation $(0 \rightarrow 8021c53a_x)$ for F_2 with a bias of $2^{-12.63}$, but no better results for F_1 and F_3[19].

Our Discovery. A longer efficient linear distinguisher could be constructed if the concatenation of forward quad-round and reverse quad-round is covered by this distinguisher. Moreover, if we use the linear approximation ($0 \rightarrow 8021c53a_x$) of F_2, we can find 24-round (from round 3 to round 26) linear approximation with a bias of $2^{-59.15}$. If the best linear approximation of F_3 ($0 \rightarrow 02400000_x$) with a bias of $2^{-13.71}$ is used, only 22-round (from round 4 to round 25) linear approximation with a bias of $2^{-51.84}$ can be discovered. If we use the best linear approximation ($0 \rightarrow 02600000_x$) of F_1 with a bias of $2^{-13.37}$, we can produce 26-round (from round 2 to round 27) linear approximation with a bias of $2^{-62.85}$ (see Fig.4). The longest found efficient linear approximations based on F_1, F_2, F_3 are listed in Table 2. In Table 2, the first column is the active round function,

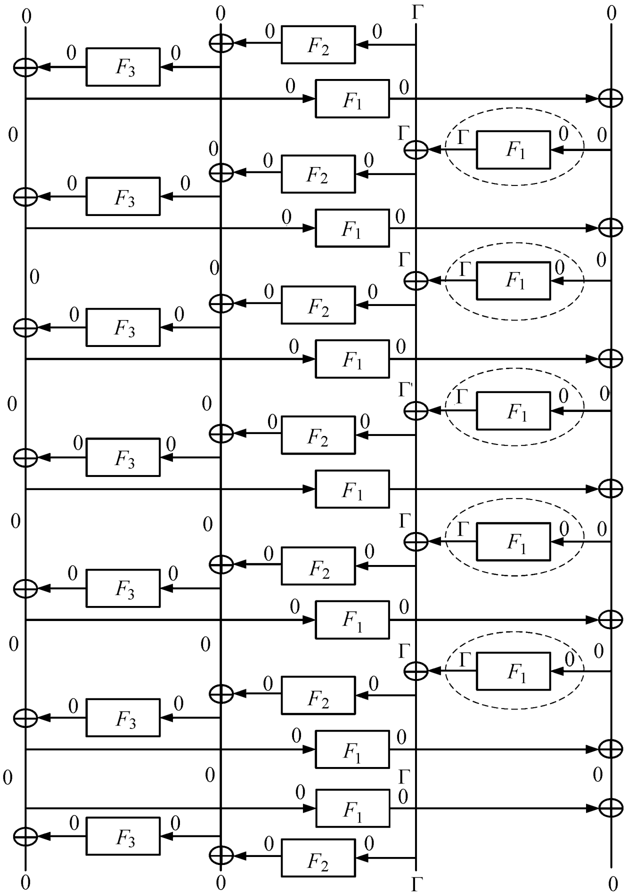

Fig.4. 26-round linear approximation of CAST-256 ($\Gamma = 02600000_x$).

Table 2. Longest Efficient Linear Approximations

F_i	Linear Approximation	Rounds (Covered Round)	Bias
F_1	$0 \xrightarrow{F_1} 02600000_x$	26 (2∼27)	$2^{-62.85}$
F_2	$0 \xrightarrow{F_2} 8021c53a_x$	24 (3∼26)	$2^{-59.15}$
F_3	$0 \xrightarrow{F_3} 02400000_x$	22 (4∼25)	$2^{-51.84}$

the second column is the best linear approximation of the corresponding active round function, the third column is the number and position of covered rounds for the identified linear approximation of CAST-256 based on the linear approximation of round function in the second column, and the last column is the bias of the identified linear approximation for CAST-256.

For example, the first row means that we use the linear approximation of F_1 to produce 26-round linear approximation for CAST-256. Linear approximation of one forward quad-round with F_1 means that only F_1 in the first round of one forward quad-round is active, and other three round functions are non-active. In this way, the bias of the linear approximation for one quad-round is equal to the bias of linear approximation of F_1 for the first round. Then we can iterate such linear approximation of one forward quad-round five times and then add one backward quad-round without the last round F_1 at the bottom of them and add one forward quad-round without the first round F_1 at the top of them, thus the total number of rounds is $3 + 4 \times 5 + 3 = 26$. For the produced linear approximation of 26 rounds of CAST-256 shown in Fig.4, the first three rounds and the last three rounds are non-active.

In Fig.4, the approximation starting from round 2 with respect to the linear approximation of the round function F_1 is depicted. The active round functions are circled, and the input and the output masks are 0 and $\Gamma = 02600000_x$, respectively. Therefore, the property of concatenation enables us to extend the linear approximation from 21 rounds to 26 rounds, which can be used to improve the linear cryptanalysis of CAST-256 significantly. The reason for this extension is that there are six non-active round functions between two active round functions F_1 at the concatenation of CAST-256. However, there are only three non-active round functions between two active round functions for other positions instead of the concatenation.

4 Key Recovery Attack of 32-Round CAST-256

4.1 Key Recovery

Using the 26-round linear approximation from round 2 to round 27 described in Fig.4, we provide a partial key recovery attack on 32-round CAST-256 which recovers the subkeys in the first round k_{r1}^1 and k_{m1}^1, the subkeys from round 28 to 32, k_{r4}^7, k_{m4}^7, k_{ri}^8 and k_{mi}^8 ($1 \leqslant i \leqslant 4$). We denote the 128-bit plaintext and ciphertext as $P = (P_1, P_2, P_3, P_4)$ and $C = (C_1, C_2, C_3, C_4)$, the input of the r-th round as (A^r, B^r, C^r, D^r), thus we have $(P_1, P_2, P_3, P_4) = (A^1, B^1, C^1, D^1)$. The key recovery attack is described in Fig.5. Assuming that N

known plaintexts are used, the partial sum technique proposed by Ferguson *et al.*[24] will be used in the partial encryption and decryption procedures. The details of the attack procedure are as follows.

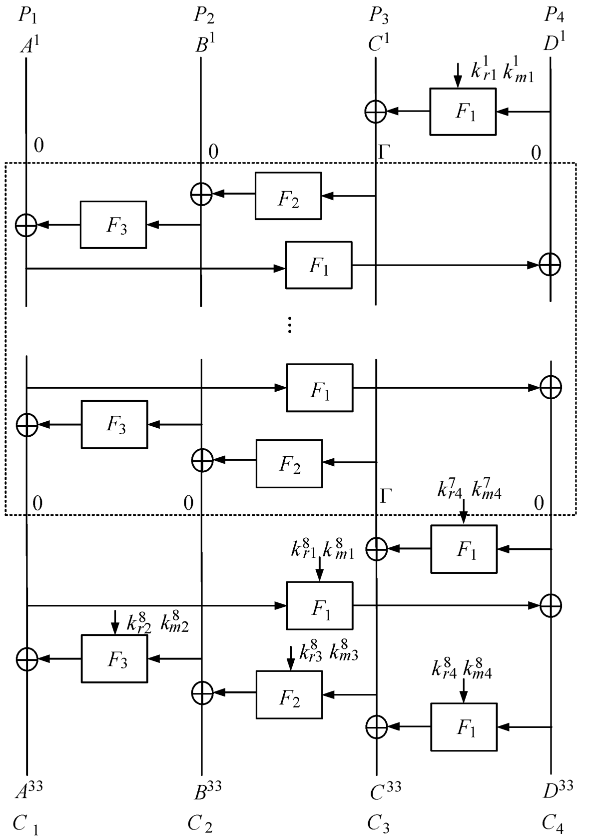

Fig.5. Key recovery of 32-round CAST-256.

1) Allocate a 32-bit counter $V_1[x_1]$ for each possible value of the 97-bit x_1: $x_1 = (D^1|(C^1 \oplus C^{30}) \cdot 02600000_x|A^{30}|D^{30})$.

2) Guess the 111-bit subkey $\{k_{m4}^8, k_{r4}^8, k_{m3}^8, k_{r3}^8, k_{m2}^8, k_{r2}^8\}$, and decrypt each ciphertext of N plaintext-ciphertext pairs for three rounds to get $x_1 = (D^1|(C^1 \oplus C^{30}) \cdot 02600000_x|A^{30}|D^{30})$, then add one to $V_1[x_1]$.

3) Allocate a 64-bit counter $V_2[x_2]$ for each possible value of the 65-bit x_2: $x_2 = (D^1|(C^1 \oplus C^{29}) \cdot 02600000_x|D^{29})$.

4) Guess the 37-bit subkey k_{m1}^8 and k_{r1}^8 in the 29th round and partially decrypt 2^{97} values for x_1, compute $x_2 = (D^1|(C^1 \oplus C^{29}) \cdot 02600000_x|D^{29})$, then add $V_1[x_1]$ to $V_2[x_2]$.

5) Allocate a 96-bit counter $V_3[x_3]$ for each possible value of the 33-bit x_3: $x_3 = (D^1|(C^1 \oplus C^{28}) \cdot 02600000_x)$.

6) Guess the 37-bit subkey k_{m4}^7 and k_{r4}^7 in the 28th round and partially decrypt 2^{65} values for x_2, compute $x_3 = (D^1|(C^1 \oplus C^{28}) \cdot 02600000_x)$, then add $V_2[x_2]$ to $V_3[x_3]$.

7) Allocate a 128-bit counter $V_4[x_4]$ for two possible values of the 1-bit x_4: $x_4 = (C^1 \cdot 02600000_x \oplus C^{28} \cdot 02600000_x)$.

8) Guess the 37-bit subkey k_{m1}^1 and k_{r1}^1 in the first round and partially encrypt 2^{33} values for x_3, compute $x_4 = (C^1 \cdot 02600000_x \oplus C^{28} \cdot 02600000_x)$, add $V_3[x_3]$ to $V_4[x_4]$. Set $\epsilon[k_{m1}^1|k_{r1}^1] = |\frac{V_4[0]}{N} - \frac{1}{2}|$.

9) After proceeding step 8, sort $\epsilon[k_{m1}^1|k_{r1}^1]$ by value in descending order. For the first 2^{31} values of $\epsilon[k_{m1}^1|k_{r1}^1]$, output the corresponding values of $(k_{m1}^1|k_{r1}^1)$ along with the guessed 185 subkey bits $(k_{r4}^7|k_{m4}^7|k_{ri}^8|k_{mi}^8)$ $(1 \leqslant i \leqslant 4)$ as candidate right subkeys③.

4.2 Estimation of Complexity

Selçuk gave the method to estimate the success probability as follows[25],

$$P_s = \Phi(2\sqrt{N} \times |p - \frac{1}{2}| - \Phi^{-1}(1 - 2^{-a-1})), \qquad (1)$$

where P_s is the success probability of attack, N is the number of known plaintexts, p is the probability that the linear approximation holds, a is the number of advantage bits we can get, Φ and Φ^{-1} are the normal distribution and its inverse, respectively.

In our attack, the probability p is $\frac{1}{2} + 2^{-62.85}$. From (1), if we set $N = 2^{126.8}$ and $a = 6$, the success probability is $P_s = 0.70$. Thus the data complexity is $2^{126.8}$ known plaintexts.

The time complexity of step 2 is $2^{111} \times N \times 3$ one-round decryptions, since we decrypt all N ciphertexts for each $2^{3\times37}$ possible key values. In step 4, 2^{97} pairs are decrypted for one round under each guess of $37 \times 4 = 148$ subkey bits, thus the time complexity in this step is $2^{4\times37} \times 2^{97} = 2^{245}$ one-round decryptions. In the similar way, the time complexity of step 6 and step 8 is $2^{5\times37} \times 2^{65} = 2^{250}$ and $2^{6\times37} \times 2^{33} = 2^{255}$ one-round decryptions or encryptions, respectively. In all, the total time complexity to recover 6-bit key information is about $2^{255} \times \frac{1}{32} = 2^{250}$ 32-round encryptions. The memory requirements are about 2^{99} bytes.

5 Conclusions

In this paper, by analyzing the property of the concatenation between forward quad-round and reverse quad-round and choosing the active round function, we constructed a 26-round linear approximation of CAST-256 which is much longer than the previous 21-round linear approximation of CAST-256. With the 26-round

③As the key schedule of CAST-256 is very complicated, we cannot recover the whole 256-bit key after we get six bits of key information. Therefore we only output $2^{222-6} = 2^{216}$ candidate right subkeys.

linear approximation, we presented a partial key information recovery attack on 32 rounds of CAST-256. The partial key information recovery attack is still significant to some extent in cryptography, therefore our attack is the best attack for CAST-256 according to the number of rounds without the weak-key assumption.

References

[1] Biham E, Shamir A. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 1991, 4(1): 3-72.

[2] Matsui M. Linear cryptanalysis method for DES cipher. In *Proc. Workshop on the Theory and Application of Cryptographic Techniques*, May 1993, pp.386-397.

[3] Knudsen L. Truncated and higher order differentials. In *Proc. the 2nd Int. Workshop on Fast Software Encryption*, December 1994, pp.196-211.

[4] Biham E, Biryukov A, Shamir A. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In *Proc. Int. Conf. the Theory and Application of Cryptographic Techniques*, May 1999, pp.12-23.

[5] Borst J, Knudsen L, Rijmen V. Two attacks on reduced IDEA. In *Proc. the 16th Advances in Cryptology-Eurocrypt*, May 1997, pp.1-13.

[6] Blondeau C, Gérard B. Multiple differential cryptanalysis: Theory and practice. In *Proc. the 18th Int. Workshop on Fast Software Encryption*, February 2011, pp.35-54.

[7] Wang M Q, Sun Y, Tischhauser E, Preneel B. A model for structure attacks, with applications to PRESENT and Serpent. In *Proc. the 19th Int. Workshop on Fast Software Encryption*, March 2012, pp.49-68.

[8] Wagner D. The boomerang attack. In *Proc. the 6th Int. Workshop on Fast Software Encryption*, March 1999, pp.156-170.

[9] Albrecht M, Cid C. Algebraic techniques in differential cryptanalysis. In *Proc. the 16th Int. Workshop on Fast Software Encryption*, February 2009, pp.193-208.

[10] Wang M, Sun Y, Mouha N, Preneel B. Algebraic techniques in differential cryptanalysis revisited. *In Proc. the 16th Information Security and Privacy Australasian Conference*, July 2011, pp.120-141.

[11] Biryukov A, De Cannière C, Quisquater M. On multiple linear approximations. In *Proc. the 24th Int. Cryptology Conf.*, August 2004, pp.1-22.

[12] Hermelin M, Cho J, Nyberg K. Multidimensional extension of Matsui's Algorithm 2. In *Proc. the 16th Int. Workshop on Fast Software Encryption*, February 2009, pp.209-227.

[13] Kaliski B, Robshaw M. Linear cryptanalysis using multiple approximations. In *Proc. the 14th Int. Cryptology Conf. Advances in Cryptology*, August 1994, pp.26-39.

[14] Bogdanov A, Rijmen V. Linear hulls with correlation zero and linear cryptanalysis of block ciphers. *Designs, Codes and Cryptography*, 2014, 70(3): 369-383.

[15] Adams C, Gilchrist J. The CAST-256 encryption algorithm, June 1999. http://www.ietf.org/rfc/rfc2612.txt, Sept. 2014.

[16] Adams C. The CAST-128 encryption algorithm, May 1997. http://www.ietf.org/rfc/rfc2144.txt, Oct. 2014.

[17] Nakahara J J, Rasmussen M. Linear analysis of reduced-round CAST-128 and CAST-256. In *Proc. the 7th Brazilian Symposium on Information and Computer System Security*, Aug. 2007, pp.45-55.

[18] Wang M Q, Wang X Y, Hu C H. New linear cryptanalytic results of reduced-round of CAST-128 and CAST-256. In *Proc. the 15th Int. Workshop on Selected Areas in Cryptography*, August 2009, pp.429-441.

[19] Sun Y, Wang M Q, Sun Q M. How to search linear approximation for large non-surjective S-box. In *Proc. the 6th ACM Symposium on Information, Computer and Communications Security*, March 2011, pp.459-465.

[20] Biham E. A note on comparing the AES candidates. In *Proc. the 2nd AES Candidate Conference*, March 1999, pp.22-23.

[21] Seki H, Kaneko T. Differential cryptanalysis of CAST-256 reduced to nine quad-rounds. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2001, 84(4): 913-918.

[22] Bogdanov A, Leander G, Nyberg K, Wang M Q. Integral and multidimensional linear distinguishers with correlation zero. In *Proc. the 18th Int. Conf. Theory and Application of Cryptology and Information Security*, December 2012, pp.244-261.

[23] Adams C M. Constructing symmetric ciphers using the CAST design procedure. *Designs, Codes and Cryptography*, 1997, 12(3): 283-316.

[24] Ferguson N, Kelsey J, Lucks S, Schneier B, Stay M, Wagner D, Whiting D. Improved cryptanalysis of Rijndael. In *Proc. the 7th Int. Workshop on Fast Software Encryption*, April 2000, pp.213-230.

[25] Selçuk A A. On probability of success in linear and differential cryptanalysis. *Journal of Cryptology*, 2008, 21(1): 131-147.

Jing-Yuan Zhao received her B.S. degree in mathematics from Shandong University, Jinan, in 2010. She is now a Ph.D. candidate of Shandong University. Her research interest is analysis of block cipher.

Mei-Qin Wang gained her B.E. and M.A. degrees from Xi'an Jiaotong University, and Ph.D. degree in information security from Shandong University, Jinan. She is a professor in School of Mathematics, Shandong University. Her research interest is the symmetric cipher.

Long Wen received his B.S. degree in mathematics from Shandong University in 2011. He is currently a Ph.D. candidate of Shandong University, Jinan. His current research interest is design and analysis of block cipher.