Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering
Shaw Conference Center, Edmonton, Alberta, Canada May 9-12 1999

An Analysis of the CAST-256 Cipher

C. Adams*, H.M. Heys†, S.E. Tavares‡, and M. Wiener*

* Entrust Technologies, 750 Heron Rd., Suite E08
Ottawa, Ontario, Canada K1V 1A7
Email: {carlisle.adams,michael.wiener}@entrust.com

† Faculty of Engineering, Memorial University of Newfoundland
St. John's, Newfoundland, Canada A1B 3X5
Email: howard@engr.mun.ca

‡ Dept. of Electrical and Computer Engineering, Queen's University
Kingston, Ontario, Canada K7L 3N6
Email: tavares@ee.queensu.ca

Abstract

In this paper, we examine the cryptographic security of the CAST-256 symmetric block encryption algorithm. The CAST-256 cipher has been proposed as a candidate for the Advanced Encryption Standard currently under consideration by the U.S. National Institute of Standards and Technology (NIST). It has been designed for a 128-bit block size and variable key sizes of up to 256 bits to suit AES requirements. In this paper, we specifically consider the cryptographic security of the cipher in relation to the cryptanalytic property of diffusion and the cryptanalysis techniques of linear and differential cryptanalysis.

1 Introduction

The CAST-256 [1] cipher is a new symmetric block cipher with a 128-bit block size and has been submitted as a candidate for the Advanced Encryption Standard (AES) [2]. The design of CAST-256 was derived from the CAST-128 cipher [3], a 64-bit block cipher, and benefits from the results of analysis of this earlier cipher [4]. Due to the relatively large block size requirement of AES, it was necessary to modify the architecture of CAST-256 from the classical Feistel structure used in CAST-128. This has an impact on the diffusion properties of the cipher and the resistance of the cipher to the typical cryptanalysis techniques applied to block ciphers. It is these characteristics of CAST-256 that we specifically consider in this paper.

The AES process, which began in 1997 by NIST, is an important development in the field of symmetric cryptography with the eventual outcome being the selection of a new block cipher to replace the aging Data Encryption Standard (DES) [5]. It is anticipated that the selected cipher will become widespread in its application, eventually becoming the standard for use in environments from banking machines to Internet email. It is critical that the selected cipher be efficient and that the general cryptographic community have confidence in its security. The AES process is currently in the first phase, with 15 accepted candidates available for public scrutiny. CAST-256 is one of these candidates.

Two of the most important cryptographic criteria of a block cipher are resistance to linear cryptanalysis [6] and resistance to differential cryptanalysis [7]. Both of these attacks have been identified in recent years as effective techniques for breaking large classes of symmetric ciphers.

2 CAST-256 Architecture

CAST-256 is an iterative product cipher consisting of 48 rounds or 12 quad-rounds of mixing, providing "confusion" and "diffusion" of data bits and key bits. The complete cipher is described thoroughly in [1]. In this section, we provide enough details of the cipher for the purposes of the discussion in this paper. Figure 1 illustrates the basic structure of one round of the CAST-256 cipher. The 128-bit input block to the

0-7803-5579-2/99/$10.00© 1999 IEEE

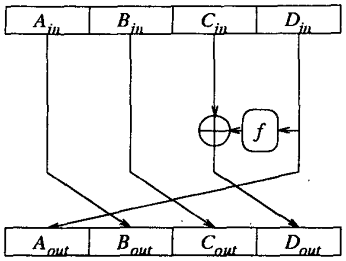

Figure 1: CAST Round Data Flow

round may be divided into four 32-bit words labelled A_{in}, B_{in}, C_{in}, and D_{in}. Word C_{in} is modified by a bit-wise exclusive-OR with the output of round function f which has word D_{in} as input. The corresponding four output words - A_{out}, B_{out}, C_{out}, and D_{out} - are then derived by rotating the words to the right by one position. This description applies specifically to the rounds in the first half of the cipher (i.e., the first 24 rounds). For the 2nd half of the cipher, the rotation in a round becomes a rotation of one word position to the left, rather than the right.

The inherent security of the CAST-256 cipher (as for all iterative block ciphers) is dependent on the round function. The CAST-256 cipher round function is directly based on the CAST-128 round function and is illustrated in Figure 2. The function S_j, $1 \leq j \leq 4$, is a nonlinear 8×32 mapping and is referred to as an S-box. Due to their nonlinear nature, the S-boxes in CAST are an integral component of cipher security.

The operations "b", "c", and "d" are referred to as *combining operations* and represent the combination of two 32-bit words. In CAST, these operations can be bit-wise exclusive-OR, addition modulo-2^{32}, and subtraction modulo-2^{32}. Operation "a" in CAST consists of combining two 32-bit words (one data, one key) using one of the 3 operations, followed by a rotation dependent on 5 bits of subkey. The round functions of CAST-256 vary between rounds, in that the combining operations used for "a", "b", "c", and "d" differ [1]. Mathematically, a *typical* round function is

$$\begin{aligned} W &= ((K_{mi} + X_i) <<< K_{ri} \\ Y_i &= ((S_1[W_1] \oplus S_2[W_2]) \\ &\quad + S_3[W_3]) - S_4[W_4] \end{aligned} \qquad (1)$$

where X_i represents the 32-bit input to the round function, S_j represents S-box j, W_j represents the 8-bit input to S-box j (the j-th byte of W), K_{mi} and K_{ri} represent the i-th round masking and rotation subkeys, respectively, Y_i represents the 32-bit output of the round function, and "$\oplus$", "+", and "$-$" represent the operations exclusive-OR, addition, and subtraction, respectively. The notation "$V <<< U$" represents a left rotation of word V by U (i.e., as determined by the least signficant 5-bits of U). Note that W, X_i, Y_i, and K_{mi} all represent 32-bit words. The vector K_{ri} is 5 bits in length. The values of subkeys K_{mi} and K_{ri}, $1 \leq i \leq 48$, are determined by a key scheduling algorithm as described in [1]. Decryption is achieved similarly to encryption, with the only changes required being the reversal of the application of the subkeys.

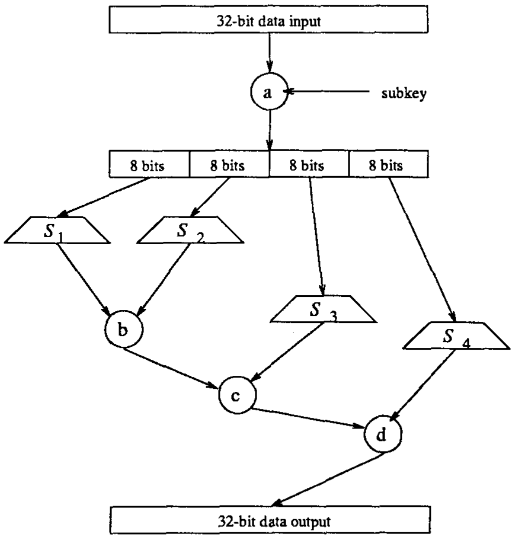

Figure 2: CAST Round Function

3 Diffusion Properties

An important cryptographic criterion of a cipher is non-degeneracy: the property that all output bits are dependent on all input bits, and vice versa. The spread of the influence of input bits to output bits is referred to as diffusion. In this section, we show that the diffusion properties of CAST-256 result in non-degeneracy in the cipher after 7 rounds (or less than 2 quad-rounds) in both the encryption and decryption directions. (We shall consider encryption only, and decryption follows similarly.)

We base our analysis on the assumptions that a round function is non-degenerate and that non-degeneracy for a bit is not cancelled through either the round function or the exclusive-OR function in the cipher architecture. Non-degeneracy in the round function is intuitively extremely likely since each S-

0-7803-5579-2/99/$10.00©1999 IEEE

Round	*Dependencies*
1	P_3 (P_4)
2	P_2 (P_3,P_4)
3	P_1 (P_2,P_3,P_4)
4	P_4 (P_1,P_2,P_3,P_4)
5	P_3 (P_1,P_2,P_3,P_4)
6	P_2 (P_1,P_2,P_3,P_4)
7	P_1 (P_1,P_2,P_3,P_4)

Table 1: Dependencies Following Cipher Rounds

box is known to be non-degenerate. We note, however, that the exclusive-OR operation is not non-degenerate as only one bit from each word influences an output bit. Hence, the critical notion in considering the diffusion properties of CAST-256 is that output word D_{out} of a round is influenced by all bits that are input to the round function (i.e., input word D_{in}) since all outputs of the round function are dependent on all inputs to the round function. However, it does not become non-degenerate in input word C_{in} and all other output words have no diffusion associated with them.

Given this, we can construct Table 1, illustrating the dependencies of the cipher data at the output of the indicated rounds. We represent the four 32-bit input words corresponding to the 128-bit plaintext block by P_1, P_2, P_3, and P_4, corresponding to the inputs A_{in}, B_{in}, C_{in}, and D_{in}, respectively, of the first round. In the *Dependencies* column, the unbracketed letter represents the block affected by the round function and the bracketed letters represent the blocks which now influence the affected block (i.e., the blocks which influence input bits to the round function).

The table indicates that after one round, the bits corresponding to the plaintext word P_3 are now non-degenerate in the bits of plaintext word P_4. Similarly, after two rounds, the word corresponding to P_2 is non-degenerate in the bits of P_3 and P_4. This contimues so that after 4 rounds (or one quad-round), the word corresponding to plaintext word P_4 is influenced by all the bits of all plaintext words. After 7 rounds the complete dependency of the output bits on the input bits has been achieved since all four words, P_1, P_2, P_3, and P_4, are influenced by all bits of the plaintext.

4 Resistance to Linear Cryptanalysis

Linear cryptanalysis [6] attempts to exploit any high-probability occurrences of linear (modulo-2) expressions of input, output, and round keys in the round function of an iterated cipher. That is, the fundamental principle of linear cryptanalysis is to find a linear approximation of the form:

$$\begin{aligned} &P_{i_1} \oplus P_{i_2} \oplus \cdots \oplus P_{i_a} \\ &\quad \oplus C_{j_1} \oplus C_{j_2} \oplus \cdots \oplus C_{j_b} \\ &\quad = K_{k_1} \oplus K_{k_1} \oplus \cdots \oplus K_{k_c} \end{aligned} \tag{2}$$

where i_1, i_2, ..., i_a, j_1, j_2, ..., j_b, and k_1, k_2, ..., k_c denote bit positions of the plaintext P, ciphertext C, and key K, respectively.

It has been estimated [6] that the best linear expression for r-rounds of a cipher has a probability of being satisfied that is bounded as follows:

$$|p_L - \frac{1}{2}| \le 2^{\alpha-1} \cdot |p_\beta - \frac{1}{2}|^\alpha \tag{3}$$

where p_L represents the probability that the linear expression (2) holds, p_β represents the probability of the best linear approximation of any S-box, and α represents the number of S-boxes involved in the linear approximation. The expression is based on the assumption of independent round keys such that the linear approximations of the S-boxes are independent. Provable immunity to linear cryptanalysis strictly depends on bounding the likelihood of an overall linear expression (sometimes referred to as a "linear hull") rather than any particular construction of a linear expression based on a specific set of S-box inputs and outputs for all rounds of the cipher. However, determining linear hull probabilities is generally an intractable problem in cipher analysis and, in this paper, we consider therefore the building block of an overall linear expression: the sequence of approximations of the round functions (involving approximations of the S-boxes) which result in the overall linear expression.

A basic linear attack typically uses a sequence of linear approximations of the rounds to create an overall linear expression involving subsets of plaintext and ciphertext bits. From this it is possible to derive the equivalent of one key bit represented as the exclusive-OR sum of a number of round key bits as shown in (2). In this case, it is shown [6] that the number of known plaintexts required is approximately

$$N_L = |p_L - \frac{1}{2}|^{-2}. \tag{4}$$

It can be shown that the best linear approximation has a probability given by

$$|p_\beta - \frac{1}{2}| = \frac{2^{m-1} - NL_{min}}{2^m} \tag{5}$$

0-7803-5579-2/99/$10.00© 1999 IEEE

where m is the number of input bits to the S-box and NL_{min} is the nonlinearity of the S-box [4]. For the S-boxes of CAST-256, $m = 8$ and $NL_{min} = 74$. Furthermore, for the CAST-256 cipher, the best linear approximation appears to be constructed by approximating one round function for every 4th round. That is, the best approximation involves 4 S-boxes every 4 rounds such that the linear approximation of the round function for every 4th round involves only output bits (i.e., the sum of some number of output bits is a constant) of the four S-boxes of the approximated round.

Since linear approximations are built using exclusive-OR, the best linear approximation of a round is based on the exclusive-OR of the one output bit of each S-box which corresponds to the least significant bit in the combining operation. For any type of combining operation (exclusive-OR, addition, or subtraction) this bit is determined by an exclusive-OR. Hence, the linear expresssion used as an approximation to the S-box is simply $Y_1 = 0$ where Y_1 represents the output bit of the S-box used as the least significant bit in the combining operation.

In the approximation of the overall cipher, we shall simplify the cipher by excluding the key-dependent rotation operation. This is equivalent to assuming that the cipher is keyed with the worst case scenario where all rotation subkey values result in a rotation of zero bits. In practice, this is highly unlikely and the rotation operation will further add to the difficulties in mounting linear crpytanalysis.

Now, based on the previous discussion, for an r-round linear approximation, $\alpha = r$. For $r = 48$, using $NL_{min} = 74$, the number of known plaintexts required in the basic linear cryptanalysis is approximately 2^{122}. Note that this is almost equal to the total number of plaintexts available (2^{128}) and argues against the practicality of a linear attack on this cipher.

Furthermore, Youssef, et al, [8] have proposed that a more accurate bound on the number of plaintexts required for linear cryptanalysis of a CAST cipher can be obtained by considering the combination of S-boxes in the round function, rather than the individual S-boxes. In particular, they compute the nonlinearity of the composite 32 × 32 S-box when the individual S-boxes are combined using exclusive-OR, an assumption only applicable to the least significant bit in the combining operations. Using this in place ofNL_{min} in the equations above and setting $m = 32$ and $\alpha = r/4$ (since we approximate the round function of every 4th round) yields a number of known plaintexts required for a 48-round linear approximation at more than 2^{176} (far beyond the number of plaintexts available).

Note that experimental evidence suggests that combining S-boxes using mixing operations such as addition or subtraction rather than exclusive-OR may increase the nonlinearity of the composite S-box even further. And this conjecture, in combination with the difficulties arising from mounting a linear attack when key-dependent rotations are used [3], provides an overwhelming argument that it is completely impractical to effectively cryptanalyze CAST-256 using a linear attack.

5 Resistance to Differential Cryptanalysis

Differential cryptanalysis [7] attempts to exploit any high probability output differences resulting from particular input differences in the round function of an iterated cipher. A block cipher can be proven to be resistant to differential cryptanalysis if it can be shown that no high probability differentials exist [9], where an r-round differential is defined to be the "difference" of two outputs after r rounds corresponding to two plaintexts with a given "difference". Generally, the most effective "differences" to consider are the bit-wise exclusive-OR of two data blocks. [1]

In a good cipher the probability of all differentials should approach 2^{-N}, where N is the block size. Strictly speaking, differential cryptanalysis requires only the existence of a highly probable differential to succeed. However, differentials can be viewed to be comprised of a number of possible characteristics, where a characteristic specifies the exact sequence of input and output exclusive-ORs for each round to achieve the overall differential input and output exclusive-OR.

It is typically difficult to derive the probability of any particular differential and, in practice, it would be hard for a cryptanalyst to determine the existence of a highly-probable differential without searching for highly-probable characteristics. Although it is often the case that an upper bound on the probability of a differential cannot be stated for a particular cipher (that is, immunity to differential cryptanalytis cannot

[1] Although it is possible to consider other differences such as the subtraction of one block from another block, there is no compelling reason to consider such differences for CAST-256. Hence, we shall focus on the more conventional exclusive-OR difference. Note that the design of CAST-256 specifically includes a mixing of the combining operations of exclusive-OR, addition, and subtraction to minimize the effectiveness of using any one particular definition of a difference.

be *proven*), the probabilities of the most likely characteristics can be estimated. These probabilities can then be used as a measure of the cipher's resistance to differential cryptanalysis.

As is common in the literature, the analysis here is based on the assumptions (1) that all round keys are independent and (2) that the occurrence of output exclusive-ORs given particular input exclusive-ORs is independent for different rounds. Under such conditions, the probability of an r-round characteristic is given by

$$p_{\Omega_r} = \prod_{i=1}^{r} p_i \qquad (6)$$

where p_i represents the probability of the output exclusive-OR given the input exclusive-OR in round i.

The best characteristics that can be constructed are typically iterative in nature. For the CAST-256 cipher with R rounds, the characteristic illustrated in Table 2 appears to be the best possible r-round iterative characteristic based on iterating a 4-round characteristic. Note that the notation $(0,0,0,\Delta)$ represents exclusive-OR vectors for the four 32-bit words in a CAST-256 round input with the first 3 words (corresponding to A_{in}, B_{in}, and C_{in} in the round input) having all zeros exclusive-OR differences and the 4th word (corresponding to D_{in} in the round input) having some non-zero exclusive-OR difference, represented by Δ. Also note that, as in the discussion for linear cryptanalysis, we assume the worst case scenario where the rotation sub-keys result in no rotations in all rounds, an extremely unlikely event.

For the characteristic illustrated for a general R-round cipher, the input exclusive-OR to round $R/2+1$ will be a vector in which one of the sub-blocks is non-zero and the other three sub-blocks are zero. (The precise variation that applies to a given cipher depends on the value of R.) Without loss of generality the example for $(0,0,0,\Delta)$ is shown in Table 2 and is suitable for the AES-defined version of CAST-256 with $R = 48$.

The input exclusive-OR and output exclusive-OR of every 4th round of the characteristic in Table 2 is of a format where a non-zero input exclusive-OR leads to an all zero output exclusive-OR (as shown for rounds 1, 5, etc.). As per the analysis and rationale given in [4], the input-output exclusive-OR pair for a simplified CAST round function (i.e., one which does not include the key-dependent rotation, and for which the only S-box combining operation used is exclusive-OR) can be assumed to have a probability of $p \le 2^{-14}$. This is based on the fact that all four S-boxes in the CAST round function are injective and the format of the exclusive-OR pair has the output exclusive-OR being equal to 0. It can be shown that it may be expected that a round function which uses combining operations such as addition and subtraction will have a reduced probability associated with the most likely input-output exclusive-OR pair of the round function [10]. (Furthermore, experimental evidence for the CAST-256 S-boxes strongly supports this analytical work.) This leads to the conclusion that the best r-round iterated characteristic, as shown in Table 2 and based on the assumptions described above, has a probability given by

$$p_{\Omega_r} \le (2^{-14})^{r/4}. \qquad (7)$$

$(0,0,0,\Delta)$	[input XOR to round 1]
$0 \leftarrow \Delta$	[round 1]
$0 \leftarrow 0$	[round 2]
$0 \leftarrow 0$	[round 3]
$0 \leftarrow 0$	[round 4]
$(0,0,0,\Delta)$	[input XOR to round 5]
$0 \leftarrow \Delta$	[round 5]
$0 \leftarrow 0$	[round 6]
$0 \leftarrow 0$	[round 7]
$0 \leftarrow 0$	[round 8]
...	repeat up to $R/2$ rounds
$(0,0,0,\Delta)$	[input XOR to round $R/2+1$]
$0 \leftarrow \Delta$	[round $R/2+1$]
$0 \leftarrow 0$	[round $R/2+2$]
$0 \leftarrow 0$	[round $R/2+3$]
$0 \leftarrow 0$	[round $R/2+4$]
...	repeat up to r rounds

Table 2: Format of Best r-round Characteristic of an R-round Cipher

In particular, a 40-round characteristic (which could potentially be used to attack the 48 round cipher) must have a probability less than or equal to 2^{-140} according to the assumptions of the analysis. This implies that the number of chosen plaintexts required for this attack would be greater than 2^{140} for the 48-round cipher (substantially greater than the number of plaintexts available for a 128-bit block size).

The results of this analysis, plus the consideration of the added difficulty in mounting a differential attack when combining operations such as addition and subtraction are used as well as exclusive-OR and key-dependent rotations are also used [3], leads to the conclusion that the CAST-256 cipher appears to be immune to differential cryptanalysis.

6 Conclusion

In this paper, we have investigated the security aspects of the CAST-256 block cipher, focussing specifically on the diffusion properties and the resistance of the cipher to linear and differential cryptanalysis. The conclusions of the analysis show that, with respect to these cipher properties, the cipher is secure. Although it is not possible to state that any cipher has guaranteed security, we believe the design philosophy of CAST-256 is sound and have shown this is the case for properties discussed in this paper.

To gain further confidence in the security of CAST-256, many other cipher properties and cryptanalysis methodologies could be analyzed for the cipher. These could include properties such as the information theoretic characteristics of the cipher and attacks such as related-key attacks, higher-order differential attacks, and linear-differential attacks. As well, the analysis of this paper could be improved upon by considering more accurately the effect of the combining operations of addition and subtraction and the use of the key-dependent rotations in each round. However, we conjecture that such a complete analysis is likely to pose intractable problems.

References

[1] C. Adams, "The CAST-256 Encryption Algorithm", available at AES web site as per reference [2], June 1998.

[2] National Institute of Standards and Technology: Advanced Encryption Standard (AES) web site: *"crsc.nist.gov/encryption/aes"*.

[3] C. Adams, "Constructing Symmetric Ciphers Using the CAST Design Procedure", *Designs, Codes, and Cryptography* vol. 12, no. 3, pp. 283-316, 1997.

[4] J. Lee, H. Heys, and S. Tavares, "Resistance of a CAST-like Encryption Algorithm to Linear and Differential Cryptanalysis", *Designs, Codes, and Cryptography*, vol. 12, no. 3, pp. 267-282, 1997.

[5] National Bureau of Standards, "Data Encryption Standard", *Federal Information Processing Standard Publication 46*, 1977.

[6] M. Matsui, "Linear Cryptanalysis Method for DES Cipher", *Advances in Cryptology: Proceedings of Eurocrypt '93*, Springer-Verlag, pp. 386-397, 1994.

[7] E. Biham and A. Shamir, *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.

[8] A. Youssef, Z. Chen, and S. Tavares, "Construction of Highly Nonlinear Injective S-boxes with Applications to CAST-like Encryption Algorithm", *Proceedings of Canadian Conference on Electrical and Computer Engineering (CCECE '97)*, St. John's, Newfoundland, May 1997, pp. 330-333.

[9] X. Lai, J. Massey , and S. Murph,y"Markov Ciphers and Differential Cryptanalysis", *Advances in Cryptology: Proceedings of Eurocrypt '91*, Springer-Verlag, pp. 17-38, 1991.

[10] L. O'Connor, "Preliminary analytical results concerning the mixing of operations from different algebraic groups and the maximum value of the resulting XOR difference distribution table", unpublished, 1998.

0-7803-5579-2/99/$10.00©1999 IEEE