On Differential Properties of Data-Dependent Rotations and Their Use in MARS and RC6

(Extended Abstract)

Scott Contini and Yiqun Lisa Yin

RSA Laboratories, 2955 Campus Drive
San Mateo, CA 94403, USA
{scontini,yiqun}@rsa.com

Abstract. Data-dependent rotations have been found to be a useful component for designing block cipher including MARS and RC6, two candidates for AES. Existing analysis indicates that data-dependent rotations play an important role in thwarting differential and linear attacks. In this paper, we study differential properties of data-dependent rotations, and derive a complete characterization of all possible characteristics for the operation. We also compare the use of data-dependent rotations in MARS and RC6.

1 Introduction

Data-dependent rotations have been found to be a useful component for designing block ciphers including MARS [3] and RC6 [11], two candidates for the Advanced Encryption Standard (AES). Implementation-wise, data-dependent rotations can be performed quickly in both software and hardware. Security-wise, they appear to be a powerful tool in preventing both differential and linear cryptanalysis.

Studies on data-dependent rotations in recent years are perhaps due to the publication of RC5 [10] which makes extensive use of data-dependent rotations. Since RC5 was proposed, various studies [2, 5, 6, 7, 8, 9, 12] have provided a greater understanding of how RC5's structure and operations, in particular data-dependent rotations, contribute to its security.

Linear approximations of data-dependent rotations have been well analyzed in the literature [7, 9, 12], and a complete characterization of all possible linear approximations for the operation was given in [9]. These results show that the mixed use of rotations and some other basic operations (e.g., addition) is a very effective way of thwarting linear cryptanalysis.

Characteristics and differentials of data-dependent rotations have also been studied, mostly in the context of differential attacks against RC5 [2, 6, 8]. All these attacks assumed that differences (in a characteristic or a differential) never occur in any "rotation amount." Such a choice is based on the heuristic argument that once a difference occurs in some rotation amount, the output difference after that rotation operation will look essentially random, and so the resulting characteristic or differential will not be useful in a differential attack. However, there was no analytic proof to justify the above heuristic.

In this paper, we focus our analysis on differential properties of data-dependent rotations. In particular, we derive a complete characterization of all possible output differences that may occur when a difference is in the rotation amount. Our main results show that the output difference is uniformly distributed over a very large set of size at least $2^{w/2}$ where w is the word size. In other words, when a difference occurs in the rotation amount, all possible characteristics for the data-dependent rotation hold with equal and very small probability. So our results provide an analytical quantification for the effectiveness of data-dependent rotations in preventing differential attacks.

We also discuss how data-dependent rotations are used in MARS and RC6. A noticeable feature of both ciphers is that the rotations are used in combination with multiplication. More specifically, a rotation amount is derived from the result of a multiplication. Such an approach ensures that any input difference to the multiplication is very likely to produce a difference in the rotation amount because of the good diffusion property of multiplication. So our results on the differential properties of rotations also provide a good justification for the combined use of data-dependent rotations and multiplication. Indeed, both MARS and RC6 seem to have strong security against differential cryptanalysis [3, 4].

2 The encryption routines in MARS and RC6

Here we give brief descriptions of the encryption routines in MARS and RC6, both of which make extensive use of data-dependent rotations. First, we introduce the notation for the basic operations that are used in the two ciphers. We use $\lg w$ to denote base 2 logarithm of w.

$a + b$	integer addition modulo 2^w
$a \oplus b$	bitwise exclusive-or of w-bit words
$a \times b$	integer multiplication modulo 2^w
$a \lll b$	rotate the w-bit word a to the left by the amount given by the least significant $\lg w$ bits of b

2.1 MARS

The cipher consists of 32 rounds of Type-3 Feistel network, divided into three phases: forward mixing, “cryptographic core,” and backward mixing. The cryptographic core of MARS has 16 iterative rounds, and each round uses a keyed E-function which is described in Table 1. Two data-dependent rotations and one multiplication are used in each E-function.

```
E-function of MARS

Input:       a 32-bit variable $in$
             two 32-bit subkeys $key1, key2$

output:      three 32-bit variables $L, R, M$

Procedure:   1. $M = in + key1$
             2. $R = (in \lll 13) \times key2$ (multiplication)
             3. $i =$ lowest 9 bits of $M$
             4. $L = S[i]$
             5. $R = R \lll 5$
             6. $M = M \lll R$ (data-dependent rotation)
             7. $L = L \oplus R$
             8. $R = R \lll 5$
             9. $L = L \oplus R$
             10. $L = L \lll R$ (data-dependent rotation)
```

Table 1. E-function of MARS

2.2 RC6

A version of RC6 is specified as RC6-$w/r/b$ where the word size is w bits, encryption consists of a nonnegative number of rounds r, and b denotes the length of the encryption key in bytes. The key schedule of RC6 expands the user supplied secret key into a set of subkeys $S[0], \ldots, S[2r+3]$. The encryption routine of RC6 consists of pre-whitening, 20 iterative rounds, and post-whitening. The round function of RC6 is given in Table 2. Two data-dependent rotations and two multiplications are used in each round.

3 Differential properties of data-dependent rotations

All published differential attacks on ciphers using data-dependent rotations assumed that differences never occur in the rotation amounts. This is quite a natural assumption, since a difference in the rotation amounts would seem to produce a random looking output difference.

In this section, we provide an analytical proof that justifies the above heuristic argument. Our main result is a complete characterization of all possible output differences that may occur when a difference is in the rotation amount. This allows us to compute precisely the probability of any characteristic for the data-dependent rotations. In particular, when the difference occurs in the rotation amounts, all possible characteristics for the data-dependent rotations hold with very small probability. As a consequence, all differentials with small Hamming

```
Round function of RC6

Input:       four $w$-bit variables $A, B, C, D$
             two $w$-bit round keys $S[2i]$, $S[2i+1]$

Output:      four $w$-bit variables $A, B, C, D$

Procedure:   1. $B = B \times (2B+1)$ (multiplication)
             2. $B = B \lll \lg w$
             3. $D = D \times (2D+1)$ (multiplication)
             4. $D = D \lll \lg w$
             5. $A = A \oplus B$
             6. $A = A \lll D$ (data-dependent rotation)
             7. $A = A + S[2i]$
             8. $C = C \oplus D$
             9. $C = C \lll B$ (data-dependent rotation)
             10. $C = C + S[2i+1]$
             11. $(A, B, C, D) = (B, C, D, A)$
```

Table 2. Round function of RC6

weight hold with very small probability. So our results quantify the effectiveness of data-dependent rotations in preventing differential attacks.

3.1 Distribution of the output difference

Given a pair of inputs (X_1, R_1) and (X_2, R_2) where X_i is a word being rotated by the value R_i, we're interested in understanding the output difference in terms of the input differences. We call the input differences X' and R' and the output difference Y'. This is summarized in the following equations.

$$\begin{aligned} Y_1 &= X_1 \lll R_1, \\ Y_2 &= X_2 \lll R_2, \\ X' &= X_1 \oplus X_2, \\ R' &= R_1 \oplus R_2, \\ Y' &= Y_1 \oplus Y_2. \end{aligned}$$

We also introduce the variable

$$r' = (R_2 - R_1) \bmod w.$$

As we will see in the analysis, the variable r' is directly related to the probability of the characteristics for data-dependent rotations. Note that "a difference is not in the rotation amount" is equivalent to $r' = 0$ or $R' \bmod w = 0$.

For a fixed input difference X', let us consider the possible output differences Y'. Since Y' appears to depend on the two rotation amounts R_1 and R_2, this motivates us to define the function

$$\begin{aligned} f_{X',R_1,R_2}(X_1) &= (X_1 \lll R_1) \oplus ((X_1 \oplus X') \lll R_2) \\ &= (X_1 \lll R_1) \oplus (X_1 \lll R_2) \oplus (X' \lll R_2) \end{aligned}$$

which, for fixed X', R_1, and R_2, expresses the output difference in terms of the input X_1. We also define

$$\begin{aligned} I_{X',R_1,R_2} &= \{Y' : Y' = f_{X',R_1,R_2}(X_1) \text{ for some } X_1\}, \\ N_{X',R_1,R_2} &= \text{size of the set } I_{X',R_1,R_2}, \\ P_{X',R_1,R_2}(Y') &= \{X_1 : f_{X',R_1,R_2}(X_1) = Y'\}. \end{aligned}$$

That is, I_{X',R_1,R_2} is the set of output differences Y' when X_1 ranges over all possible values and $P_{X',R_1,R_2}(Y')$ is the set of inputs X_1 which yield output difference Y'. The letters I and P stand for image and pre-image, respectively.

Theorem 1. *(1)* $N_{X',R_1,R_2} = 2^{w-\gcd(w,r')}$.
(2) For any $Y' \in I_{X',R_1,R_2}$*, the size of the set* $P_{X',R_1,R_2}(Y')$ *is* $2^{\gcd(w,r')}$.

Before proving the theorem, we first discuss some of its implications by contrasting the case where $r' = 0$ with the case where $r' \neq 0$:

1. $r' = 0$. The difference is not in the rotation amount.
   In this case, we have $\gcd(w, r') = w$ and $N_{X',R_1,R_2} = 1$. In other words, there is only *one* possible output difference Y'. All the characteristics used in existing differential attacks on RC5, RC6 and MARS belong to this category.
2. $r' \neq 0$. The difference is in the rotation amount.
   In this case, $gcd(w, r')$ is a power of 2 between 1 and $w/2$. Hence, N_{X',R_1,R_2} ranges between $2^{\frac{w}{2}}$ and 2^{w-1}, and each possible output different difference occurs *exactly* the same number of times. In other words, the output difference Y' is uniformly distributed in a set of size at least $2^{\frac{w}{2}}$ when the pair of inputs with a fixed difference ranges over all possible values.

From the above comparison, we can see that a difference in the rotation amount is spread out in the output difference in a drastic way.

We now move on to the proof of Theorem 1. We will recall some facts from group theory to simplify our understanding of the set of output differences I_{X',R_1,R_2}. The set of $w-$bit words form a group isomorphic to Z_2^w under the operation of exclusive-or. For a fixed integer r, the function $h(X) = X \lll r$ is a homomorphism: it has the property that $h(X \oplus Y) = h(X) \oplus h(Y)$. The function $p(X) = \text{parity}(X)$ is a homomorphism from Z_2^w to Z_2, and the kernel of p is the subgroup of even parity words, isomorphic to Z_2^{w-1}. Exclusive-oring any odd parity word to this subgroup yields the coset of odd parity words.

Proof of Theorem 1: We have $I_{X',R_1,R_2} = \{(X_1 \lll R_1) \oplus (X_1 \lll R_2) \oplus (X' \lll R_2)\}$. By replacing X_1 with the same value rotated right by R_1, we get a more convenient definition of the set:

$$I_{X',R_1,R_2} = \{X_1 \oplus (X_1 \lll r') \oplus (X' \lll R_2)\}.$$

Since $(X' \lll R_2)$ is a constant for fixed X' and R_2, the structure of I_{X',R_1,R_2} is determined by

$$g(X_1) = X_1 \oplus (X_1 \lll r'). \tag{1}$$

Let

$$S = \{g(X_1) : \ X_1 \text{ is a } w\text{-bit word}\}.$$

It is easy to verify that g is a homomorphism from the group of $w-$bit words (a group isomorphic to Z_2^w) to S, and S is a subgroup.

First, consider the special case where r' is odd. We claim that in this case S is isomorphic to Z_2^{w-1}. To prove this, we only need to show that the kernel of g has exactly two elements. The kernel consists of the X_1's satisfying

$$X_1 = X_1 \lll r'.$$

This property implies conditions on the bits of X_1: bit 0 must be the same as bit r', bit 1 must be the same as bit $r' + 1 \bmod w$, and so on. Since r' is relatively prime to w, we have that all bits must be the same, showing that the kernel is the two elements containing all 0's and all 1's.

Therefore, S is a subgroup isomorphic to Z_2^{w-1}. In particular, it is the subgroup of even parity words, and I_{X',R_1,R_2} is the coset of words having the same parity as $(X' \lll R_2)$. Hence $N_{X',R_1,R_2} = 2^{w-1} = 2^{w-\gcd(w,r')}$. The second part of the theorem follows from elementary group theory.

For the general case, we write $r' = 2^i r$ where r is odd. Then we can show that $2^{\gcd(w,r')} = 2^{2^i}$ is the size of the kernel of g, where the kernel elements are of the form $a|a|\ldots|a$ and a can be any 2^{2^i} values for a 2^i-bit vector (there are $w/2^i$ a's concatenated). □

3.2 Characteristics

Based on Theorem 1, we can easily compute the probability of any characteristic for the data-dependent rotations when the difference is in the rotation. For a binary vector X, we will use $|X|$ to denote the *Hamming weight* of X.

Theorem 2. *For $i = 1, 2$, let $Y_i = X_i \lll R_i$. Let $r' = (R_2 - R_1) \bmod w$. Then each characteristic holds with either probability 0 or probability $2^{\gcd(w,r')-w}$.*

Note that for $w = 32$, the above theorem implies that the probability of any characteristic is at most $2^{-w/2} = 2^{-16}$ when the difference occurs in the rotation amount.

Corollary 3. *For $i = 1, 2$, let $Y_i = X_i \lll R_i$. Let $r' = (R_2 - R_1) \bmod w$. If $|X'| = 0$, then the probability that $|Y'| = 0$ is $2^{\gcd(w,r')-w}$.*

The following two corollaries follow from the proof of Theorem 1, since the parity of Y' must be the same as the parity of X'.

Corollary 4. *For $i = 1, 2$, let $Y_i = X_i \lll R_i$. Let $r' = (R_2 - R_1) \bmod w$. If $|X'| = 0$, then the probability that $|Y'| = 1$ is 0.*

Corollary 5. *For $i = 1, 2$, let $Y_i = X_i \lll R_i$. Let $r' = (R_2 - R_1) \bmod w$. If $|X'| = 1$, then the probability that $|Y'| = 0$ is 0.*

3.3 Differentials with small Hamming weights

From Theorem 2, we know that when the difference is in the rotation amount, all possible characteristics for data-dependent rotations hold with equal and very small probability. We now turn our attention towards differentials for the data-dependent rotations. This section will focus on differentials of Hamming weight one. The analysis can be extended to more general differentials.

Theorem 6. *For $i = 1, 2$, let $Y_i = X_i \lll R_i$. Let $r' = (R_2 - R_1) \bmod w$ and write $r' = 2^i r$ where r is odd. For a given input difference X' such that $|X'| = 1$, the probability that $|Y'| = 1$ is $p = \frac{1}{2^{w - \lg(w) - 2^i + i}}$.*

Proof. The probability of the differential is

$$p = \frac{\text{size of the set } \{Y' : Y' \in I_{X', R_1, R_2} \text{ and } |Y'| = 1\}}{N_{X', R_1, R_2} = 2^{w - 2^i}}. \tag{2}$$

To evaluate the numerator, we rewrite Y' using the definition of $g(X_1)$ given in Equation 1.

$$Y' = g(X_1) \oplus (X' \lll R_2).$$

Since $|X'| = 1$, there are only two possibilities for $|g(X_1)|$ in order to have $|Y'| = 1$. That is, (1) $|g(X_1)| = 0$, and (2) $|g(X_2)| = 2$, and one of the two 1-bits in $g(X_1)$ lines up with the 1-bit in X'.

We claim that the set of values of $g(X_1)$ with Hamming weight 2 are those words in which the two 1-bits are a multiple of 2^i positions apart from each other. It is not difficult to see that any word with Hamming weight 2 in which the two 1's are a multiple of 2^i positions apart from each other belongs to the set S. The more challenging part is showing that these are the *only* Hamming weight two words within the set S. We start by considering the "first" such word of this form: $1 + 2^i$. This word, along with all rotations of it, generate the subgroup S, which has size $2^{w - 2^i}$. In fact, one basis for S consists of this word rotated left by j positions for $j = 0$ to $w - 2^i - 1$. $g(x)$ cannot contain a Hamming weight 2 word where the 1's are a distance 2^{i-1} apart (for example), since that would cause it to generate a subgroup of size $2^{w - 2^{i-1}}$ which is larger than the subgroup under consideration.

Completing the proof, the size of the set $\{g(X_1) : |g(X_1)| = 2 \text{ and } |Y'| = 1\}$ is $\frac{w}{2^i} - 1$, since one of 1-bits in $g(X_1)$ lines up with the 1-bit in X' and there are exactly $\frac{w}{2^i} - 1$ positions for the other bit "1". Adding on the one case where $g(x) = 0$, the numerator of equation 2 becomes $\frac{w}{2^i}$ and hence the probability is $\frac{1}{2^{w - \lg(w) - 2^i + i}}$. □

Corollary 7. *Let the word size $w = 32$. If $|X'| = 1$ and there is a difference in the rotate amounts, then the probability that $|Y'| = 1$ is $\leq 2^{-15}$.*

Similar analysis shows that, for data-dependent rotations, all differentials with small Hamming weight hold with very small probability if there is a difference in the rotation amount. Therefore, it seems very unlikely that such differentials could be helpful to improve existing differential attacks on RC5, RC6, and MARS.

4 The use of data-dependent rotations in MARS and RC6

In this section we discuss the use of data-dependent rotations in MARS and RC6. Before doing so, we first consider how the data-dependent rotation is used in RC5, which actually provides some good insight into the design of RC6 and MARS.

Perhaps two of the most distinguishing features of RC5 are the heavy use of data-dependent rotations and exceptional simplicity in its design. While no practical attack on RC5 has been found, existing studies provide some theoretical differential style attacks, generally based on the fact that the rotation amounts in RC5 only depend on the few least significant bits in a word. This particular problem in using data-dependent rotations have motivated researchers to find better ways of using data-dependent rotations that can take full advantages of the operation.

MARS and RC6 are indeed two such examples. Interestingly, the approach that is used in both ciphers to "strengthen" data-dependent rotations is somewhat similar. In particular, the approach is to combine data-dependent rotations with integer multiplication.

Let us now take a closer look at how the rotation amounts are derived in MARS and RC6. There two data-dependent rotations in each round of MARS, and also two in each round of RC6. For ease of comparison, we let r_1 and r_2 denote the two rotation amounts in each round of both ciphers. We use $R[s..t]$ to denote bits s through t of R.

In MARS, r_1 and r_2 are computed from the result of multiplication between an intermediate value and a subkey. More precisely,

$$\begin{aligned} R &= in \times key2, \\ r_1 &= R[31..27], \\ r_2 &= R[26..22]. \end{aligned}$$

In RC6, r_1 and r_2 are computed from the result of two data-dependent multiplications, respectively, where the data are some intermediate values. More precisely,

$$B = B \times (2B + 1),$$

$$
\begin{aligned}
D &= D \times (2D+1)),\\
r_1 &= B[31..27],\\
r_2 &= D[31..27].
\end{aligned}
$$

To summarize, MARS uses a *keyed* linear function (of the data), while RC6 uses a *keyless* quadratic function (of the data). Even though the way how multiplication is introduced is quite different in the two ciphers, they both take advantage of the good diffusion property of multiplication. In particular, the high order bits of the output from multiplication are used as subsequent rotation amount. For both the linear function in MARS and the quadratic function in RC6, it has been shown [3, 4] that any input difference in the data will result a difference in one of the high order bits of the output with high probability. Based on our results on differential properties of data-dependent rotations, we can conclude that combining multiplication with data-dependent rotation is a very effective way of preventing differential attacks.

References

1. E. Biham and A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard.* Springer-Verlag, New York, 1993.
2. A. Biryukov and E. Kushilevitz. Improved cryptanalysis of RC5. In K. Nyberg, editor, *Advances in Cryptology — Eurocrypt '98*, volume 1403 *Lecture Notes in Computer Science*, pages 85–99, 1998. Springer Verlag.
3. C. Burwick, D. Coppersmith, E. D'Avignon, R. Gennaro, S. Halevi, C. Jutla, S. Matyas Jr., L. O'Connor, M. 'Peyravian, d. Safford, and N. Zuric. MARS - a candidate cipher for AES. IBM Corporation, June 10, 1998.
4. S. Contini, R.L. Rivest, M.J.B. Robshaw and Y.L. Yin. The Security of the RC6 Block Cipher. v1.0, August 20, 1998. Available at `www.rsa.com/rsalabs/aes/`.
5. M.H. Heys. Linearly weak keys of RC5. *IEE Electronic Letters*, Vol. 33, pages 836–838, 1997.
6. B.S. Kaliski and Y.L. Yin. On differential and linear cryptanalysis of the RC5 encryption algorithm. In D. Coppersmith, editor, *Advances in Cryptology — Crypto '95*, volume 963 of *Lecture Notes in Computer Science*, pages 171–184, 1995. Springer Verlag.
7. B.S. Kaliski and Y.L. Yin. On the Security of the RC5 Encryption Algorithm. RSA Laboratories Technical Report TR-602. Available at `www.rsa.com/rsalabs/aes/`.
8. L.R. Knudsen and W. Meier. Improved differential attacks on RC5. In N. Koblitz, editor, *Advances in Cryptology — Crypto '96*, volume 1109 of *Lecture Notes in Computer Science*, pages 216–228, 1996. Springer Verlag.
9. S. Moriai, K. Aoki, and K. Ohta. Key-dependency of linear probability of RC5. March 1996. To appear in *IEICE Trans. Fundamentals.*
10. R.L. Rivest. The RC5 encryption algorithm. In B. Preneel, editor, *Fast Software Encryption*, volume 1008 of *Lecture Notes in Computer Science*, pages 86–96, 1995. Springer Verlag.
11. R.L. Rivest, M.J.B. Robshaw R. Sidney and Y.L. Yin. The RC6 Block Cipher. v1.1, August 20, 1998. Available at `www.rsa.com/rsalabs/aes/`.

12. A. A. Selcuk. New results in linear cryptanalysis of RC5. In S. Vaudenay, editor, *Fast Software Encryption*, volume 1372 of *Lecture Notes in Computer Science*, pages 1–16, 1998, Springer-Verlag.

This article was processed using the LaTeX macro package with LLNCS style