Improved Analysis of
Some Simplified Variants of RC6

Scott Contini[1], Ronald L. Rivest[2], M.J.B. Robshaw[1], and Yiqun Lisa Yin[1]

[1] RSA Laboratories, 2955 Campus Drive
San Mateo, CA 94403, USA
{scontini,matt,yiqun}@rsa.com
[2] M.I.T. Laboratory for Computer Science, 545 Technology Square
Cambridge, MA 02139, USA
rivest@theory.lcs.mit.edu

Abstract. RC6 has been submitted as a candidate for the Advanced Encryption Standard (AES). Two important features of RC6 that were absent from its predecessor RC5 are a *quadratic function* and a *fixed rotation*. By examining simplified variants that omit these features we clarify their essential contribution to the overall security of RC6.

1 Introduction

RC6 is an evolutionary improvement of the block cipher RC5 [9] that was designed to meet the requirements of the Advanced Encryption Standard (AES). Like RC5, RC6 makes essential use of data-dependent rotations, but it also includes new features such as the use of four working registers instead of two, and the inclusion of integer multiplication as an additional primitive operation. Two components of RC6 that were absent from RC5 are a *quadratic function* to mix bits in a word more effectively and a *fixed rotation* that is used both to hinder the construction of good differentials and linear approximations and also to ensure that subsequent data dependent rotation amounts are more likely to be affected by any ongoing avalanche of change.

An initial analysis of the security of RC6 and its resistance to the basic forms of differential and linear cryptanalysis was given in [3]. Here we further illustrate how these new operations contribute to the security of RC6 by studying simplified variants (that is, intentionally weakened forms) of RC6. In particular, our approach is to find the best attack on the weakened forms and then try to adapt the attack to the full cipher. Since one of the design principles of RC6 was to build on the experience gained with RC5, the focus of our analysis will be in assessing the relevance to RC6 of the best existing cryptanalytic attacks on RC5. We will often refer to the work of Knudsen and Meier [8] and that of Biryukov and Kushilevitz [2]. These authors in particular have made very significant advances in understanding the security of RC5.

Our work splits naturally into two parts. The first focuses on the usefulness of the fixed rotation and the second on the quadratic function. While our analysis is targeted at RC6 and its simplified variants, some of the results might well be

of independent interest. Our analysis starts by considering some of the weakened variants of RC6 that were introduced in [3]. More specifically, by dropping the fixed rotation we derive a cipher that we will denote by RC6-NFR (where NFR stands for no fixed rotation), by dropping the quadratic function we obtain RC6-I (where I stands for the identity function), and by dropping both operations we have RC6-I-NFR.

We will consider characteristics and differentials for RC6-I-NFR and RC6-NFR that have already been described in [3]. We study the relations between certain values of the subkeys and the probability of a characteristic and/or differential. Such phenomena are similar to the "differentially-weak keys" of RC5 observed by Knudsen and Meier [8]. We describe our observations and provide a thorough analysis which suggests that inclusion of the fixed rotation destroys the structure required for such dependencies to form. As a consequence RC6-I and RC6 itself seem to be immune from any direct extension of the results previously obtained on RC5.

Second, we examine the diffusive properties of the quadratic function and other operations that are used in RC6. In this analysis we track the Hamming weight (the number of 1's) of the exclusive-or difference between two quantities as they are encrypted. Quite naturally this leads to the idea of differentials that are constructed using such a measure of difference and this notion is very similar in spirit to earlier work on RC5 [2, 8]. We show that the quadratic function drastically increases the Hamming weight of some input difference when the Hamming weight of an input difference is small. This indicates that the use of both the quadratic function and data-dependent rotations in RC6 make it unlikely that differential attacks similar to those that were useful for RC5 [2, 8] can be effectively extended to RC6.

2   Description of RC6 and variants

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. Throughout this paper we will set $w = 32$, $r = 20$, $b = 16$, 24, or 32 and we will use RC6 to refer to this particular version. The base-two logarithm of w will be denoted by $\lg w$ and RC6 uses the following six basic operations:

$a + b$	integer addition modulo 2^w
$a - b$	integer subtraction 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
$a \ggg b$	rotate the w-bit word a to the right by the amount given by the least significant $\lg w$ bits of b

The user supplies a key of length k bytes which is then expanded to a set of subkeys. The key schedule of RC6 is described in [10]. Since here we are

only concerned with encryption, we will assume that the subkeys $S[0]$, ..., $S[43]$ are independent and chosen at random. RC6 works with four w-bit registers A, B, C, D which contain the initial input plaintext as well as the output ciphertext at the end of encryption. We use $(A, B, C, D) = (B, C, D, A)$ to mean the parallel assignment of values on the right to registers on the left.

Encryption with RC6-w/20/b

Input: Plaintext stored in four w-bit input registers A, B, C, D
        w-bit round keys $S[0, \ldots, 43]$

Output: Ciphertext stored in A, B, C, D

Procedure: $B = B + S[0]$
            $D = D + S[1]$
            for $i = 1$ to 20 do
                {
                    $t = (B \times (2B + 1)) \lll \lg w$
                    $u = (D \times (2D + 1)) \lll \lg w$
                    $A = ((A \oplus t) \lll u) + S[2i]$
                    $C = ((C \oplus u) \lll t) + S[2i + 1]$
                    $(A, B, C, D) = (B, C, D, A)$
                }
            $A = A + S[42]$
            $C = C + S[43]$

The three simplified variants of RC6 that we will consider throughout the paper are distinguished from RC6 in the way the values of t and u are assigned. These differences are summarized in the following table.

The assignment of t and u in RC6 and some weakened variants			
	RC6-I-NFR	*RC6-I*	*RC6-NFR*	*RC6*
$t =$	B	$B \lll \lg w$	$B \times (2B + 1)$	$(B \times (2B + 1)) \lll \lg w$
$u =$	D	$D \lll \lg w$	$D \times (2D + 1)$	$(D \times (2D + 1)) \lll \lg w$

3   The fixed rotation

In [8] Knudsen and Meier show that the values of some of the subkeys in RC5 can have a direct effect on the probability of whether some differential holds. In this section we show that a similar phenomenon can be observed in weakened variants of RC6 that do not use the fixed rotation. This should perhaps come as little surprise since while the structure of RC6-I-NFR is very different to that of RC5, it uses the same operations and might be expected to have similar behavior at times. We will then consider the role of the fixed rotation used in RC6 and we will demonstrate by analysis and experimentation that the effects seen in RC5 and some simplified variants of RC6 do not seem to exist within RC6 itself.

3.1 Existing analysis on RC6-I-NFR and RC6-NFR

In [3] one potentially useful six-round iterative characteristic was provided for attacking both RC6-I-NFR and RC6-NFR. This is given in Table 1. Here e_t is used to denote the 32-bit word that has all bits set to zero except bit t where $t = 0$ for the least significant bit. We use A_i (respectively B_i, C_i and D_i) to denote the values of registers A (respectively B, C, and D) at the beginning of round i. As an example, A_1, B_1, C_1, and D_1 contain the plaintext input after pre-whitening and for the six-round variants of the cipher, A_7, B_7, C_7 and D_7 contain the output prior to post-whitening. According to [3], when averaged over all possible subkeys, the expected probability that this characteristic holds is 2^{-30} for both RC6-I-NFR and RC6-I.

3.2 Refined analysis of RC6-I-NFR and RC6-NFR

Closer analysis of the characteristic probabilities for RC6-I-NFR and RC6-NFR suggests that the values of some of the subkeys during encryption are important. In particular, the characteristic of interest for RC6-I-NFR and RC6-NFR given in Table 1 can only occur if certain subkey conditions are met. Further, once these subkey conditions hold then the characteristic occurs with probability 2^{-20}, which is much higher than the initial estimate of 2^{-30} that was obtained by averaging over all subkeys.

i	A_i	B_i	C_i	D_i
1	e_{31}	e_{31}	0	0
		$\downarrow$		
2	e_{31}	0	0	0
		$\downarrow$		
3	0	0	0	e_{31}
		$\downarrow$		
4	0	e_{31}	e_{31}	0
		$\downarrow$		
5	e_{31}	e_{31}	0	e_{31}
		$\downarrow$		
6	e_{31}	e_{31}	e_{31}	0
		$\downarrow$		
7	e_{31}	e_{31}	0	0

Table 1. A characteristic for RC6-I-NFR and RC6-NFR.

In the analysis that follows we will concentrate on RC6-NFR. The same arguments and results can be applied to RC6-I-NFR by replacing $f(x) = x \times (2x + 1)$ with the identity function $f(x) = x$. We will use the fact that $x \bmod 2^i$

uniquely determines $(x \times (2x+1)) \bmod 2^i$. Furthermore, the notation "$=_{32}$" will be used to indicate when two values are congruent modulo 32.

Lemma 1. *If the characteristic given in Table 1 holds for RC6-NFR, then the following two conditions on the subkeys must hold:*

$$f(-S[9]) =_{32} -S[7],$$
$$f(S[8]) =_{32} -S[11].$$

Proof. First we observe that if the characteristic is to hold, then certain rotation amounts derived from the B and D registers must be zero. Note that we always have that $B_i = A_{i+1}$ and that $D_i = C_{i+1}$. As a consequence, for the characteristic to hold we must have

$$D_2 =_{32} C_3 =_{32} 0, \qquad B_3 =_{32} A_4 =_{32} 0,$$
$$B_4 =_{32} A_5 =_{32} 0, \qquad D_4 =_{32} C_5 =_{32} 0,$$
$$B_5 =_{32} A_6 =_{32} 0, \qquad B_6 =_{32} A_7 =_{32} 0.$$

Using the fact that the rotation amounts are 0, we get the following two equations from rounds three and four and rounds four and five.

$$B_4 = (C_3 \oplus f(D_3)) + S[7], \tag{1}$$
$$B_5 = (C_4 \oplus f(D_4)) + S[9]. \tag{2}$$

Since $B_4 =_{32} 0$, $C_3 =_{32} 0$, $B_5 =_{32} 0$ and $D_4 =_{32} 0$, we have $S[7] =_{32} -f(D_3)$ and $C_4 =_{32} -S[9]$. Since $C_4 = D_3$, we obtain the first condition on subkeys $S[7] =_{32} -f(-S[9])$.

Similarly, looking at the computation from rounds four and five and rounds five and six, we get the following two equations.

$$D_5 = A_4 \oplus f(B_4) + S[8], \tag{3}$$
$$B_6 = C_5 \oplus f(D_5) + S[11]. \tag{4}$$

Since $A_4 =_{32} 0$, $B_4 =_{32} 0$, $B_6 =_{32} 0$ and $C_5 =_{32} 0$, we have $D_5 =_{32} S[8]$ and $S[11] =_{32} -f(D_5)$, and so $S[11] =_{32} -f(S[8])$. $\qquad\square$

The subkey dependencies in Lemma 1 were obtained using only four equations (those for B_4, B_5, D_5 and B_6). In total, one could write down 12 equations of the form $B_{i+1} = (((C_i \oplus f(D_i)) \lll f(B_i)) + S[2i+1]$ and $D_{i+1} = (((A_i \oplus f(B_i)) \lll f(D_i)) + S[2i]$ for this characteristic. Although there might be dependencies involving other equations, the four given above will be the focus of the rest of this section. Essentially, each equation involves four variables and the aim is to combine equations to obtain two expressions with a single variable. If the two expressions involve the same variable then we can obtain conditions on the subkeys involved. The four equations we use are the only ones from the set of twelve that allow us to do this.

It is worth noting that given such conditions on the subkeys involved not only does the characteristic hold, but it does so with a higher probability than the expected value given in [3].

Lemma 2. *Assume that the characteristic given in Table 1 holds up to round five. Furthermore suppose that $f(-S[9]) =_{32} -S[7]$ and $f(S[8]) =_{32} -S[11]$. Then $B_5 =_{32} 0$ and $B_6 =_{32} 0$.*

Proof. From Lemma 1, we have that $S[7] =_{32} -f(D_3)$. This is equivalent to $-S[7] =_{32} f(C_4)$. Also, we have that $B_5 =_{32} C_4 + S[9]$. So, if $-S[7] =_{32} f(-S[9])$ then $f(C_4) =_{32} f(-S[9])$ which implies that $C_4 =_{32} -S[9]$ and so $B_5 =_{32} 0$. A similar argument can be used to show that $B_6 =_{32} 0$. $\square$

Lemma 2 shows that when the subkey conditions hold, $B_5 =_{32} 0$ and $B_6 =_{32} 0$. In this case the probability of the characteristic will be $2^{-30} \times 2^5 \times 2^5 = 2^{-20}$, since two of the rotation amounts are always zero. Recall that the estimated probability for the characteristic when averaged over all keys is 2^{-30} [3]. Here we have shown (Lemmas 1 and 2) that there is some irregularity in the distribution of the probability: For a fraction of 2^{-10} keys the probability is 2^{-20}, and for the rest of the keys the probability is much smaller than 2^{-30}. This kind of irregular distribution can sometimes be exploited as was demonstrated by Knudsen and Meier with RC5 [8] who showed some techniques for using it in a differential attack. We would expect the same to apply here. Similar subkey dependencies can be observed for some of the other characteristics for RC6-I-NFR and RC6-NFR given in [3]. However in some cases the characteristic must be iterated more than once before dependencies exist.

Note that the behavior of the differential associated with some characteristic is typically of more importance in a differential attack. For RC6-I-NFR, while the characteristic displays the irregular behavior already described, the associated differential has been experimentally verified to hold with the expected probability [3]. However the associated differential for RC6-NFR appears to have the same irregular behavior as the characteristic. Why is there this discrepancy? In [3] it is shown how the introduction of the quadratic function helps to reduce the additional effect of differentials. In short, for RC6-I-NFR there are many equally viable paths that match the beginning and end-points of the characteristic. If the characteristic fails to hold because of some choice of subkey values, other characteristics hold instead thereby maintaining the probability of the differential. However, with RC6-NFR we introduce the quadratic function and this typically reduces differentials to being dominated by the action of a single characteristic. Irregular behavior in the characteristic will therefore manifest itself as irregular behavior in the differential.

3.3   Differential characteristics in RC6-I and RC6

Let us now consider the role of the fixed rotation that was omitted in RC6-I-NFR and RC6-NFR. We will find that this single operation removes the kind of subkey dependencies that occurred in these two variants.

We will focus on RC6-I in the analysis for simplicity, and the same arguments also apply to the full RC6. We will need to make some heuristic assumptions to make headway with our analysis. Nevertheless our experimental results confirm

that the differential behavior of RC6-I is pretty much as expected. It also closely matches the behavior described in [3].

Consider the characteristic given in Table 2. This is the characteristic which seemed to be one of the most useful for attacking RC6-I [3]. We first argue that there are no subkey dependencies of the form we described in Section 3.2 for this characteristic and we then broaden our discussion to include other, more general, characteristics.

i	A_i	B_i	C_i	D_i
1	e_{16}	e_{11}	0	0
		$\downarrow$		
2	e_{11}	0	0	0
		$\downarrow$		
3	0	0	0	e_{26}
		$\downarrow$		
4	0	e_{26}	e_{26}	0
		$\downarrow$		
5	e_{26}	e_{21}	0	e_{26}
		$\downarrow$		
6	e_{21}	e_{16}	e_{26}	0
		$\downarrow$		
7	e_{16}	e_{11}	0	0

Table 2. A useful characteristic for RC6-I.

At this stage we need some new notation and the exponent n will be used to denote when some quantity has been rotated to the left by n bit positions. For example, $D_2^5 =_{32} 15$ means that when D_2 is rotated five bits to the left, then the decimal value of the least significant five bits is 15. Of course, this is the same as saying that the most significant five bits of D_2 take the value 15.

For simplicity, we will assume that $(x + y)^j = x^j + y^j$ where j denotes a rotation amount. This is true if, and only if, there is no carry-out when adding the top j bits and no carry-out when adding the bottom $32 - j$ bits. For the sake of our analysis however we make this assumption, since it should actually facilitate the construction of any potential subkey dependencies!

Following the arguments in Lemma 1, for the characteristic in Table 2 to hold the following rotation amounts must take the values indicated:

$$D_2^5 =_{32} C_3^5 =_{32} 15, \qquad B_3^5 =_{32} A_4^5 =_{32} 27,$$
$$B_4^5 =_{32} A_5^5 =_{32} 27, \qquad D_4^5 =_{32} C_5^5 =_{32} 27,$$
$$B_5^5 =_{32} A_6^5 =_{32} 17, \qquad B_6^5 =_{32} A_7^5 =_{32} 17.$$

We wish to write down four equations similar to Equations (1), (2), (3) and (4) which cause subkey dependencies in RC6-NFR. From round three to four,

the difference e_{26} is copied from register D_3, is changed to e_{31} by the action of the fixed rotation, and then exclusive-ored into the C strand. For it to become the e_{26} that appears in B_4, the data dependent rotation B_3^5 must have the value 27. Hence, we must have $B_3^5 =_{32} 27$ and $B_4 = (C_3 \oplus D_3^5)^{27} + S[7] = C_3^{27} \oplus D_3 + S[7]$. In a similar way other equations can be derived:

$$B_4 = C_3^{27} \oplus D_3 + S[7], \tag{5}$$

$$B_5 = C_4^{27} \oplus D_4 + S[9], \tag{6}$$

$$D_5 = A_4^{27} \oplus B_4 + S[8], \tag{7}$$

$$B_6 = C_5^{17} \oplus D_5^{22} + S[11]. \tag{8}$$

In Lemma 1 we observed a subkey dependency by combining the analogous equations to (5) and (6), and another dependency from combining the analogous equations to (7) and (8). In the case of RC6-I we can demonstrate that neither approach now works.

We first consider Equations (5) and (6). For Equation (6) we know that the values of $B_5^5 \bmod 32$, $D_4^5 \bmod 32$, and $S[9]^5 \bmod 32$ are fixed. This implies a condition on the least significant five bits of C_4. Since C_4 is the same as D_3, we have a condition on $D_3 \bmod 32$. We now have conditions on all the registers in Equation (5), namely, $B_4^5 \bmod 32$, $C_3^5 \bmod 32$, and $D_3 \bmod 32$. However the bits from different words involved in this equation are from different positions. They don't lead to any constraints on $S[9]$, and there appear to be no subkey dependencies as a result.

Similarly arguments also apply to Equations (7) and (8). One may also try to combine Equations (5) and (7), since they have the quantity B_4 in common, or Equations (6) and (8), since they have $C_5 = D_4$ in common. However, these combinations once again fail to give any subkey dependencies.

We performed experiments on RC6-I to assess the probability of the characteristics given in Table 2. These results confirmed that the distribution of the characteristic probability was as expected, and there was no indication of any subkey dependencies for the characteristic.

More generally, we might consider characteristics of the form given in Table 3. The values which we need to fix if the characteristic is going to hold are

$$D_2^5 =_{32} C_3^5 =_{32} s - t, \qquad B_3^5 =_{32} A_4^5 =_{32} u - 5 - s,$$
$$B_4^5 =_{32} A_5^5 =_{32} u - 5 - s, \qquad D_4^5 =_{32} C_5^5 =_{32} v - u - 5,$$
$$B_5^5 =_{32} A_6^5 =_{32} u - 15 - v, \qquad B_6^5 =_{32} A_7^5 =_{32} u - 15 - v.$$

Let $r_1 = u - 5 - s$, $r_2 = v - u - 5$, and $r_3 = u - 15 - v$. Then the subkey dependencies we observed would be produced by the following equations:

$$B_4 = C_3^{r_1} \oplus D_3^{5+r_1} + S[7],$$
$$B_5 = C_4^{r_1} \oplus D_4^{5+r_1} + S[9],$$
$$D_5 = A_4^{r_2} \oplus B_4^{5+r_2} + S[8],$$
$$B_6 = C_5^{r_3} \oplus D_5^{5+r_3} + S[11].$$

i	A_i	B_i	C_i	D_i
1	e_{t+5}	e_t	0	0
		$\downarrow$		
2	e_t	0	0	0
		$\downarrow$		
3	0	0	0	e_s
		$\downarrow$		
4	0	e_u	e_s	0
		$\downarrow$		
5	e_u	e_{u-5}	0	e_v
		$\downarrow$		
6	e_{u-5}	e_{u-10}	e_v	0
		$\downarrow$		
7	e_{u-10}	e_{u-15}	0	0

Table 3. A generalized characteristic for RC6-I.

Following similar arguments to those presented earlier, it can be verified that there is no choice for r_1, r_2, and r_3 that makes the characteristic depend upon the values of the subkeys. In particular, the most promising values to try are $r_1 = 0$; $r_1 = 27$; $r_3 = 0$ and $r_2 = 22$; and $r_3 = 0$, $r_2 = 27$, and $r_1 = 27$.

The fixed rotation is an important component of RC6. Not only does it help to hinder the construction of good differentials and linear approximations [3] but it helps to disturb the build-up of any inter-round dependencies. Here the fixed rotation ensures that equations can simultaneously hold without forcing any restriction on the values of the quantities involved.

4   The quadratic function

In this section, we examine the diffusive properties of the quadratic function and other operations used in RC6. Both the work of Knudsen and Meier [8] and that of Biryukov and Kushilevitz [2] rely on the following fact about RC5: It has a relatively slow avalanche of change from one round to the next, unless the difference in two words is in the bits used to determine a data-dependent rotation. When that happens, the amount of change in one round to the other can be dramatic, but until then the rate of change tends to be rather modest. This can be exploited to a limited degree in attacks on RC5 [2, 8].

We will choose a measure of diffusion that complements naturally the work given in [2, 8]. We will use the Hamming weight of the exclusive-or difference between two words as a measure of the difference, rather than the actual value of the difference as we would in differential cryptanalysis [1] or part of the difference as we would in truncated differential cryptanalysis [7]. It is straightforward to

envisage using this notion of difference in a differential-style attack, something we call *Hamming weight differentials*, and this is very similar to some of the earlier analysis of RC5 [2, 8]. While this earlier work focused on how to effectively use such differentials to attack RC5, the focus of our work will be on assessing the likely impact of the quadratic function in thwarting such attacks.

Even for a simple operation it can be difficult to fully characterize the probability distribution of the Hamming weight of some output difference given the Hamming weight of the input differences. We will study the problem by analyzing the *expected* Hamming weight of such an output difference and it turns out that such an approach provides a good insight into the role of the different operations.

Our analysis shows that the quadratic function drastically increases the Hamming weight of some difference especially when the Hamming weight of the input difference is small. This illustrates a nice effect whereby the use of the quadratic function complements that of the data-dependent rotation. As we have mentioned, the data-dependent rotation becomes an effective agent of change only when there is a difference in the rotation amount. With a small Hamming weight difference, it is less likely that non-zero difference bits appear in positions that affect a rotation amount. However, the quadratic function helps to drastically increase the avalanche of change so that the full benefit of the data-dependent rotations can be gained as soon as possible.

4.1 Definitions and assumptions

We introduce some useful notation and definitions. For a w-bit binary vector X, let $|X|$ denote the *Hamming weight* of X, i.e., $|X|$ is the number of 1's in X. Throughout this paper we will be continually referring to RC6 and so we will assume that the word size $w = 32$. We will let $X' = X_1 \oplus X_2$, $Y' = Y_1 \oplus Y_2$, and $Z' = Z_1 \oplus Z_2$ and we use x, y, z to denote the Hamming weight of the differences $|X'|, |Y'|, |Z'|$, respectively.

Let us consider the following two conditions that may be imposed on some difference that has Hamming weight x.

A: There is a single block of consecutive 1's of length x, and the block is distributed randomly at some position in the input difference.

B: There are $t > 1$ blocks of consecutive 1's of length $x_1, x_2, ..., x_t$ such that $x_1 + x_2 + \cdots + x_t = x$. In addition, each block is distributed randomly across the input difference.

Condition B is actually a good characterization for the differences in the intermediate rounds of RC6 and its variants. In each round (of RC6 or its variants) any difference in the A and C strands are rotated by a random amount due to the data-dependent rotations. Hence each block of 1's within the differences is distributed randomly. Condition A is a special case of Condition B. In the next two sections when we examine the diffusive properties of individual operations, we will first consider the special case Condition A and then generalize the results to Condition B.

4.2 Diffusive properties of the basic operations

Here we analyze the basic operations of exclusive-or, addition, and rotation. The more complicated quadratic function will be considered in the next section.

Lemma 3. (exclusive-or) *For $i = 1, 2$ let $Z_i = X_i \oplus Y_i$. If X' and Y' satisfy Condition A, then $E(z) = x + y - \frac{2xy}{w}$.*

Proof. Since the block of 1's in X' and Y' is distributed randomly, each bit "1" in X' overlaps with each bit "1" in Y' with probability $\frac{1}{w}$. So the expected length of overlap in the output difference is $\frac{xy}{w}$, implying that the expected Hamming weight of the output is $x + y - \frac{2xy}{w}$. ☐

Corollary 4. (exclusive-or) *For $i = 1, 2$ let $Z_i = X_i \oplus Y_i$. If X' and Y' satisfy Condition B then $E(z) = x + y - \frac{2xy}{w}$.*

Proof. Follows directly from the proof of Lemma 3. ☐

Note that the expected overlap between the quantities X' and Y' is similar to the number of "corrections" used by Biryukov and Kushilevitz in their analysis of *corrected Fibonacci sequences* [2]. There an explicit formula was not provided [2] but all sequences with a "reasonable" number of corrections were experimentally generated and this was used as an estimate in their work.

Lemma 5. (addition) *For $i = 1, 2$ let $Z_i = X_i + S$, where S is the subkey. If X' satisfies Condition A then averaging over all possible X_1, X_2, S, $E(z) = c + \frac{x+1}{2}$ where $c \in [0, 1]$ and depends on X'.*

Proof. We start with the special case where $|X'| = w$, that is, X_1 and X_2 differ in all bits. We first prove that when averaging over all possible X_1, X_2, S,

$$prob(X_1 + S < 2^w \text{ and } X_2 + S \geq 2^w) = \frac{1}{4}. \tag{9}$$

Given any $X_1 \in \{0, 1\}^w$, we define

$$d(X_1) = |S : S \in \{0, 1\}^w, \text{s.t. } X_1 + S < 2^w \text{ and } X_2 + S \geq 2^w|.$$

If $X_1 < 2^{w-1}$, we have $d(X_1) = X_2 - X_1 = (X_1 \oplus (2^w - 1)) - X_1 = 2^w - 1 - 2X_1$. (If $X_1 \geq 2^{w-1}$, $d(X_1) = 0$.) Hence,

$$prob(X_1 + S < 2^w \text{ and } X_2 + S \geq 2^w) = \frac{\sum_{X_1=0}^{2^{w-1}-1} d(X_1)}{2^w \times 2^w} = \frac{1}{4}.$$

Note that for Equation 9 the particular value of w is unimportant. So we can consider the least significant j bits of X_1, X_2, S. More precisely, for $1 \leq j \leq w$, define $X_1(j) = X_1 \bmod 2^j, X_2(j) = X_2 \bmod 2^j, S(j) = S \bmod 2^j$. Then,

$$prob(X_1(j) + S(j) < 2^j \text{ and } X_2(j) + S(j) \geq 2^j) = \frac{1}{4}. \tag{10}$$

By symmetry,

$$prob(X_1(j) + S(j) \geq 2^j \text{ and } X_2(j) + S(j) < 2^j) = \frac{1}{4}. \qquad (11)$$

From Equations 10 and 11, we know that with probability $1/2$, exactly one of the two addition operations ($X_1 + S$ and $X_2 + S$) produces a carry into bit j. If this happens, Z_1 and Z_2 will be the same in bit j. Therefore, with probability $1/2$, the jth bit ($j \geq 1$) of $Z' = Z_1 \oplus Z_2$ is 1. Since bit 0 of Z' is always 1, the expected Hamming weight of Z' is $\frac{w-1}{2} + 1 = \frac{w+1}{2} = c + \frac{w+1}{2}$ for $c = 0$. We have proved the Lemma for the special case where $|X'| = w$.

Let us now consider the general case where $|X'| = x$ for some $1 \leq x \leq w$. Let v be the index of the most significant 1 in X'. So X_1 and X_2 are the same in bits $v + 1$ through $w - 1$. When computing $Z_1 = X_1 + S$ and $Z_2 = X_2 + S$, it is possible that one or both of the carries will propagate into bits $v + 1$ and higher. It is not hard to show that the "extra" number of bit differences between Z_1 and Z_2 due to this carry effect has an expectation c for some $0 \leq c \leq 1$. So the expected Hamming weight of the output difference is $c + \frac{x+1}{2}$. □

Corollary 6. (addition) *For $i = 1, 2$ let $Z_i = X_i + S$, where S is the subkey. Suppose that X' satisfies Condition B and there are t blocks of 1's in X'. Then averaging over all possible keys S, $E(z) \leq t + \frac{x+t}{2}$.*

Proof. Follows from Lemma 5. □

The fixed rotation $Z = X \lll \lg w$ always preserves the Hamming weight of the input difference in the output difference. For data-dependent rotations, it is straightforward to see that provided the input difference does not affect the rotation amount, then the Hamming weight of the difference is preserved. We can state this simple fact in the following lemma.

Lemma 7. (data-dependent rotation) *For $i = 1, 2$ let $Z_i = X_i \lll Y_i$. If $Y' =_w 0$, then $z = x$.*

The more interesting case is when $Y' \neq_w 0$. It has previously been shown [4, 6] that once a difference in the amount of rotation is experienced then the output difference is distributed in an essentially random manner over a very large set. This essentially makes any differential-style attack impossible since in this case there is a very substantial diffusive effect. So depending on the difference Y', a data-dependent rotation can either preserve the Hamming weight or increase the Hamming weight by a significant amount. The probability of the latter case occurring is closely related to the Hamming weight of Y' and we have the following lemma that characterizes such a relation for the special case.

Lemma 8. *Let $y = |Y'|$ and let p be the probability that $Y' \neq_w 0$. If Y' satisfies Condition A then $p = \min\left(\frac{y + \lg w - 1}{w}, 1\right)$.*

For the more general case when Y' satisfies Condition B it is not so simple to derive a precise formula similar to the one given above. However it is clearly the case that the heavier the Hamming weight of Y', the larger the probability that some part of the non-zero input difference will have an effect on the rotation amount.

4.3 Diffusive properties of the quadratic function

Here we consider the diffusive properties of the quadratic function $Z = f(X)$, an important new operation in RC6. First, we restate a lemma regarding the quadratic function that first appeared in [3]. This lemma characterizes the behavior of the output when a single bit of some input is flipped.

Lemma 9. [3] *Given an input X_1 chosen uniformly at random from $\{0,1\}^{32}$, let $g_{i,j}$ denote the probability that flipping bit i of X_1 will flip bit j of $Z_1 = f(X_1)$. Then,*

$$g_{i,j} = \begin{cases} 0 \; for \; j < i, \\ 1 \; for \; j = i, \\ 1 \; for \; j = 1 \; and \; i = 0, \; and \end{cases}$$
$$g_{i,j} \in [1/4, 3/4] \; for \; j > i \geq 1 \; or \; j \geq 2 \; and \; i = 0.$$

For the last case, $g_{i,j}$ is close to $3/4$ if $j = 2i + 2$, and for most of the other i, j pairs $g_{i,j}$ is close to $1/2$.

Put descriptively this lemma shows that flipping bit i of some input X will always flip bit i of the output and will, in most cases, also flip bit j where $j > i$ of the output with probability around $1/2$.

We can extend the lemma to the more general case where multiple bits of the input are flipped and we obtain a similar result: Let i be the bit position of the least significant 1 in X'. Then flipping bit i of the input X_1 will always flip bit i of the output and will, in most cases, flip bit j for $j > i$ of the output with probability around $1/2$. Experiments confirm both this intuition and also the following, perhaps surprising, result.

Lemma 10. (quadratic function) *For $i = 1, 2$ let $Z_i = f(X_i)$. Let $x = |X'|$ and $z = |Z'|$. If X' satisfies Condition A then $E(z) \approx 1 + \frac{x+w-2}{4}$.*

Proof. Let i be the index of the least significant 1 in X'. For a fixed i, the expected value of z is roughly $1 + (w - 1 - i)/2$. If X' satisfies Condition A then i is uniformly distributed between 0 and $(w - x)$. Hence,

$$E(z) \approx \frac{1}{(w - x) + 1} \sum_{i=0}^{w-x} \left(1 + \frac{w - 1 - i}{2}\right) = 1 + \frac{x + w - 2}{4}.$$

$\square$

Corollary 11. (quadratic function) *For $i = 1, 2$ let $Z_i = f(X_i)$. Let $x = |X'|$ and $z = |Z'|$. If X' satisfies Condition B and there are t blocks of 1's in X', then $E(z) \geq 1 + \frac{x+w+t-3}{4}$.*

Proof. Similar to the proof of Lemma 10. □

Lemma 10 shows that even when the difference in some input to the quadratic function has Hamming weight 1, the average Hamming weight of the difference in the output is 8.75. This is a very important result. All the other basic operations in RC6, as well as those used in RC5, generally provide little or no additional change to the output difference if the Hamming weight of the input difference is very low.

We can illustrate the effect of including the quadratic function in the following way. We experimentally measure the probability that the rotation amounts[3] at the end of a given number of rounds are unaffected by a single bit change in the first word of the input to the cipher. We consider rotation amounts in this exercise because current differential-style attacks on RC5 and RC6 require any difference propagating through the cipher to leave the rotation amounts unchanged. We use "-" to indicate that experimentally the probability is approximately (2^{-20}), which is indistinguishable from random noise.

Rounds	RC6-I-NFR	RC6-I	RC6-NFR	RC6
2	$2^{-0.54}$	$2^{-0.64}$	$2^{-1.32}$	$2^{-10.27}$
4	$2^{-2.15}$	$2^{-2.45}$	$2^{-6.27}$	-
6	$2^{-6.14}$	$2^{-7.04}$	$2^{-14.30}$	-
8	$2^{-12.76}$	$2^{-14.97}$	-	-
10	$2^{-19.07}$	-	-	-

For an increased number of rounds, the probability of unchanged rotation amounts gives a good illustration of the relative diffusive effect of RC6 and its weakened variants. It also illustrates the role of the quadratic function in the security of RC6.

Basic differential-style attacks attempt to predict and control the change from one round to the next during encryption [5]. Improved attacks on RC5 [2, 8] do not attempt to predict the difference quite so closely. Instead, they rely on the relatively slow diffusive effect of RC5 to ensure that any change propagating through the cipher remains manageable and to some extent predictable. Even though single-bit starting differences might be used, differentials with an ending difference of Hamming weight 15, for example, can still be useful [2, 8].

The quadratic function was added to RC6 to address this particular short-coming of RC5 and our work suggests that the quadratic function is likely to hinder attacks that rely on a modest avalanche of change from one round to the next.

---

[3] By "rotation amounts" we mean the low five bits of the registers for RC6-I-NFR and RC6-NFR, the high five bits of the registers for RC6-I, and the high five bits of the output of $f(x)$ for RC6.

5   Conclusions

In this paper we have considered the role of two operations in RC6 that differentiate it from RC5. Both operations are essential to the security of RC6. It is interesting to observe that RC6-I-NFR, a simplified variant of RC6 without either of these operations, has some of the behavior of RC5. RC6-I-NFR tends to have a slow rate of diffusion thereby potentially providing opportunities to mount differential attacks similar to those described for RC5 [2, 8]. Further, RC6-I-NFR demonstrates some of the differentially-weak key phenomena that has also been observed in RC5 [8]. The introduction of both the fixed rotation and the quadratic function makes RC6 resistant to such shortcomings.

We stress the importance of simplicity when designing a cipher. Unnecessary complexity makes it hard to perform a systematic examination of the true security offered. By contrast, the exceptional simplicity of RC5 invites others to assess its security. This tradition continues with RC6 with a design that encourages the researcher and aims to facilitate a deep understanding of the cipher.

Acknowledgements

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. 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/`.
4. S. Contini and Y.L. Yin. On differential properties of data-dependent rotations and their use in MARS and RC6. *To appear.*
5. 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.
6. 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/`.
7. L.R. Knudsen. Applications of higher order differentials and partial differentials. In B. Preneel, editor, *Fast Software Encryption*, volume 1008 of *Lecture Notes in Computer Science*, pages 196–211, 1995. Springer Verlag.
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. 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.
10. 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/`.