Analysis of RC2

March 24, 2003

Executive Summary

This report contains an analysis of the block cipher RC2.

No serious flaws or weaknesses have been identified in the design which could lead to practical cryptanalytic attacks with respect to the state-of-the-art and seems to provide reasonable security. However, RC2 is not a fast cipher.

The differential attack outlined in this report can find the secret key on input about 2^{60} chosen plaintexts independent of the key length. Thus, for key sizes bigger than 64, there is an attack on RC2 which, in theory, is faster than an exhaustive search for the key. Although this attack is of limited practical use, it illustrates that RC2 needs the number of rounds specified by the authors. Since RC2 is not a very fast cipher, one could be tempted to decrease the number of rounds, but this is not recommended.

Finally we mention that this report is the result of a limited time of review. A longer, concentrated analysis might reveal properties of RC2 which we were not able to detect.

Contents

1 Introduction

RC2 is a 64-bit block cipher with keys of variable lengths. It is designed by Ron Rivest for RSA Data Security, Inc., and is used widely, most notably in the S/MIME secure messaging standard [11]. In 1997 RC2 was published as an Internet Draft[34].

The model for cryptanalysis in this report is the standard one for block ciphers, where it is assumed that the attacker has access to a black-box which encrypts any chosen input or decrypts any chosen output using RC2 with a secret, randomly chosen key.

2 Structural features and characteristics

RC2 works on 64-bit blocks which are divided into four words of each sixteen bits. It is an iterated block cipher where the ciphertext is computed as a function of the plaintext and the secret key in a number of rounds. There are two kinds of rounds in RC2, the mixing rounds and the mashing rounds. There are in total 16 mixing rounds and two mashing rounds. In each round each of the four sixteen-bit words in an intermediate ciphertexts is updated as a function of the other words. Each of the mixing rounds takes a 16-bit subkey. The 64 subkeys are derived from the user-selected key which can be of length from one to 128 bytes. An additional parameter of the algorithm is the effective key length, which will be explained below.

We note that the decryption operation does not equal the encryption operation which may have unfortunate impacts on implementations.

Also, RC2 is not a fast cipher and an optimized version of DES and any of the five AES [30] finalists is likely to produce higher throughputs than RC2.

2.1 Key Expansion

The key-schedule takes a user-selected key and a number representing the maximum effective key length. The latter is a feature not seen in any other block ciphers as far as this author is informed. Assume that the user-selected key consists of T bytes where $1 \leq T \leq 128$. Let L be a key buffer (an array) of 128 bytes. The T bytes are loaded into $L[0]$, ..., $L[T-1]$ of the key buffer. The maximum effective key length in bits is denoted $T1$. The effective key length in bytes $T8$ is defined as $T8 = \lceil T1/8 \rceil$ and a mask TM as $TM = 255 \bmod 2^{8(1-T8)+T1}$. As a first observation we found that the latter is equivalent to $TM = 2^{T1 \bmod 8} - 1$.

The key expansion consists of the following two iterations, where Π is a table consisting of a permutation of the numbers $0, \ldots, 127$ derived from the expansion of π:

```
1. for $i = T, T+1, ..., 127$ do
       $L[i] = \Pi[L[i-1] + L[i-T]]$, where addition is modulo 256
2. $L[128 - T8] = \Pi[L[128 - T8] \,\&\, TM]$
3. for $i = 127 - T8, ..., 0$ do
       $L[i] = \Pi[L[i+1] \oplus L[i+T8]]$
```

Finally, define the 64 subkeys, $K[i]$ for $i = 0, \ldots, 63$ as follows: $K[i] = L[2i] + 256 \times L[2i+1]$.

The terms $T8$ and TM ensure that the expanded key table is derived from only $T1$ bits, such that an exhaustive search can be performed in 2^{T1} operations independent of the length of the user-selected key.

2.2 Encryption and Decryption

The two kinds of rounds in RC2 are defined via the operations `MIX` and `MASH`. The plaintext is divided into four words of each sixteen bits denoted $R[0], \ldots, R[3]$.

The MIX operation is defined as follows, where $s[0] = 1$, $s[1] = 2$, $s[2] = 3$, and $s[3] = 5$.

$$\begin{aligned}
&R[i] = R[i] + K[j] + (R[i-1] \mathbin{\&} R[i-2]) + ({\sim}R[i-1] \mathbin{\&} R[i-3]);\\
&j = j + 1;\\
&R[i] = R[i] \lll s[i];
\end{aligned}$$

Here the indices of R are computed modulo 4, and j is a pointer such that $K[j]$ is the first key word in the expanded key which has not yet been used in a MIX operation.

A mixing round consists of four consecutive MIX steps, such that each of the words $R[0], R[1], R[2]$, and $R[3]$ are modified and in that order.

The MASH operation is defined as follows:

$$R[i] \quad = \quad R[i] + K[\,R[i-1] \mathbin{\&} 003f_x\,],$$

in other words, the least significant six bits of $R[i-1]$ are used to select one of the 64 subkeys.

A mashing round consists of four MASH operations such that each of the words $R[0], R[1], R[2]$, and $R[3]$ are modified.

1. Let the words $R[0]$, ..., $R[3]$ hold the 64-bit plaintext block.
2. Perform the key expansion such that the words $K[0], \ldots, K[63]$ hold the subkeys.
3. Initialize j to zero.
4. Do five mixing rounds.
5. Do one mashing round.
6. Do six mixing rounds.
7. Do one mashing round.
8. Do five mixing rounds.
9. The ciphertext is defined as the resulting values of $R[0]$, ..., $R[3]$.

Decryption is the reverse of encryption. Clearly, it suffices to define the inverse operations of the MIX and MASH operations. The inverses of the MIX operations are defined as follows:

$$\begin{aligned}
&R[i] = R[i] \ggg s[i];\\
&R[i] = R[i] - K[j] - (R[i-1] \mathbin{\&} R[i-2]) - ({\sim}R[i-1] \mathbin{\&} R[i-3]);\\
&j = j - 1;
\end{aligned}$$

The inverses of the MASH operations are defined as follows:

$$R[i] \quad = \quad R[i] - K[\,R[i-1] \mathbin{\&} \mathtt{003f_x}\,].$$

An inverse mixing round consists of four consecutive inverse MIX steps, such that each of the words $R[3], R[2], R[1]$, and $R[0]$ are modified and in that order, and similarly for the inverse mashing rounds. Decryption can now be defined.

1. Let the words $R[0], \ldots, R[3]$ hold the 64-bit ciphertext block.

2. Perform the key expansion such that the words $K[0], \ldots, K[63]$ hold the subkeys.
3. Initialize j to 63.
4. Do five inverse mixing rounds.
5. Do one inverse mashing round.
6. Do six inverse mixing rounds.
7. Do one inverse mashing round.
8. Do five inverse mixing rounds.
9. The plaintext is defined as the resulting values of $R[0], \ldots, R[3]$.

3 Differential cryptanalysis

In [21] a detailed differential analysis of RC2 was published. Since this author was deeply involved in this work, the analysis given here is greatly inspired by the work of [21].

In this analysis of RC2 the difference between two 16-bit words A and B is defined to be $A \oplus B$. For ciphers which mixes modular additions with binary operations it has proved advantageous to consider differences of low Hamming weights. The reason for this is that characteristics involving multiple-bit exclusive-or differences over integer addition will generally hold with lower probability than single-bit characteristics [13].

Let e_t denote a 16-bit word with a single one bit in position t from the right, all other bits being set to zero. Also, let the leftmost bit of a 16-bit word be the most significant bit. As an example, e_{15} denotes a 16-bit word with the only non-zero bit being the most significant bit. The word of 16 zero bits will be denoted as 0_x where the subscript x denotes hexadecimal notation. And the Hamming weight, that is, the number of ones in the binary expansion of some quantity x, is denoted as $\mathrm{Hwt}(x)$.

For the remainder of the paper, we shall consider mixing and mashing rounds in the following way. Instead of viewing the operation at each step as acting on a different word we shall consider the operations to be identical, that is, at each `MIX` step

$$R[0] = R[0] + K[j] + (R[3] \,\&\, R[2]) + ({\sim}R[3] \,\&\, R[1])),$$

but that between steps the words are rotated cyclically, that is,

$$\texttt{TEMP} = R[0]; R[0] = R[1]; R[1] = R[2]; R[2] = R[3]; R[3] = \texttt{TEMP}.$$

Let us begin by considering characteristic for the `MIX` steps. Given an input difference $(e_t, 0_x, 0_x, 0_x)$ to the first `MIX` step in a mixing round, the output difference before rotation will be $(e_t, 0_x, 0_x, 0_x)$ with probability $p \geq 1/2$. The rotation then moves this single bit difference within the particular word, and the four words are swapped cyclically. There are four basic characteristics which hold with probability $p \geq 1/2$ (when averaged over all plaintexts and key words) for a `MIX` step. The value of the rotation $s[i]$ depends on the step i in which the characteristic is applied, cf. earlier. Note that addition within the subscript of e_t is to be performed modulo 16.

$$(e_t, 0_x, 0_x, 0_x) \rightarrow (0_x, 0_x, 0_x, e_{t+s[i]}) \quad (1)$$

$$(0_x, 0_x, 0_x, e_t) \rightarrow (0_x, 0_x, e_t, 0_x) \quad (2)$$

$$(0_x, 0_x, e_t, 0_x) \rightarrow (0_x, e_t, 0_x, 0_x) \quad (3)$$

$$(0_x, e_t, 0_x, 0_x) \rightarrow (e_t, 0_x, 0_x, 0_x) \quad (4)$$

With $t = 15$ (1) holds with probability $p = 1$, the three other characteristics hold with probability $p = 1/2$ on the average, which can easily be checked. In a chosen plaintext attack the attacker can choose the plaintexts carefully such that the first one-round characteristics will hold with high probability. The following examples illustrate this point.

- In (2), if $(R[2] \;\&\; e_t) = (R[1] \;\&\; e_t)$ then $p = 1$.
- In (3), if $(R[3] \;\&\; e_t) = 0_x$ then $p = 1$.
- In (4), if $(R[3] \;\&\; e_t) = e_t$ then $p = 1$.

Let us next consider characteristics for the `MASH` steps. The basic step is

$$R[0] = R[0] + K[\,R[3]\,\&\,003f_x].$$

Given an input difference $(0_x, 0_x, 0_x, e_t)$ to a mashing round with $(e_t \;\&\; 003f_x) = 0_x$ the same key word $K[\cdot]$ will be added to both sets of partially encrypted data. There are four basic useful characteristics for `MASH`:

$$\begin{aligned}
(e_t, 0_x, 0_x, 0_x) &\quad\rightarrow\quad (0_x, 0_x, 0_x, e_t) &\quad (5)\\
(0_x, 0_x, 0_x, e_t) &\quad\rightarrow\quad (0_x, 0_x, e_t, 0_x) &\quad (6)\\
(0_x, 0_x, e_t, 0_x) &\quad\rightarrow\quad (0_x, e_t, 0_x, 0_x) &\quad (7)\\
(0_x, e_t, 0_x, 0_x) &\quad\rightarrow\quad (e_t, 0_x, 0_x, 0_x) &\quad (8)
\end{aligned}$$

Characteristic (5) holds with probability $p = 1$ when $t = 15$ and otherwise $p = 1/2$, characteristics (7) and (8) hold with probability $p = 1$, and characteristic (6) holds with probability $p = 1$ if $(e_t \;\&\; 0x3f) = 0_x$. Joining these four characteristics together to pass across a mashing round with probability $p = 1$ is straightforward.

Next we attempt to combine characteristics for both the mixing and the mashing rounds of RC2. To facilitate an analysis it will be assumed that the subkey words $K[0], \ldots, K[63]$ are independent. The subkey words are not independent in RC2, but our tests (see later) reveal that this is a plausible assumption to make for differential attacks. Also, nobody has so far been able to incorporate subkey dependencies in differential cryptanalysis. The aim of the attacks here is to recover the expanded key table $K[\cdot]$.

The characteristics of greatest interest are those built from differences of low Hamming weights, and as noted earlier there are advantages to having a single non-zero bit in the most significant bit of a word. Because of the different rotation amounts in the `MIX` steps in a mixing round this leads to some conditions on t, the position of the single-bit difference. Another advantageous approach is to build characteristics such that the mashing rounds are passed with probability one. If a one-bit characteristic specifies an input difference to a mashing round of e_t in any one of the words, then provided $t = 15$ the characteristic will pass through the mashing round unhindered with probability $p = 1$. If $6 \leq t < 15$, the best characteristic has probability $p = 1/2$, because of the modular addition.

The observations provided so far allow us to present in Table 1 the differentials that are useful to us.

A more accurate assessment of the success of a differential attack can be obtained by considering "differentials" [24] instead of characteristics. In the following the issue of differentials is considered in more detail.

A characteristic describes one particular path of differences through the encryption rounds. From one given difference there might well be other "paths" through the cipher to the same target difference than the one described by one particular

plaintext difference	difference after 15 rounds	prob.	possible values of t
$(e_t, 0, 0, 0)$	$(e_{t+15}, 0, 0, 0)$	2^{-58}	4
$(e_t, 0, 0, 0)$	$(e_{t+15}, 0, 0, 0)$	2^{-59}	1, 2, 3
$(0, e_t, 0, 0)$	$(0, e_{t+14}, 0, 0)$	2^{-58}	5
$(0, e_t, 0, 0)$	$(0, e_{t+14}, 0, 0)$	2^{-59}	1, 3
$(0, e_t, 0, 0)$	$(0, e_{t+14}, 0, 0)$	2^{-60}	0, 2, 4
$(0, 0, e_t, 0)$	$(0, 0, e_{t+13}, 0)$	2^{-58}	14
$(0, 0, e_t, 0)$	$(0, 0, e_{t+13}, 0)$	2^{-59}	7, . . ., 13
$(0, 0, 0, e_t)$	$(0, 0, 0, e_{t+11})$	2^{-58}	6
$(0, 0, 0, e_t)$	$(0, 0, 0, e_{t+11})$	2^{-59}	15, 0, . . ., 5

Table 1: 26 15-round characteristics that are useful in a differential attack on RC2 and which all pass the mashing rounds with probability $p \geq 1/2$.

characteristic. The probability of the differential is given by the sum of the probabilities of all the characteristics that satisfy the differential. For RC2 it turns out that there is some “differential effect” in the characteristics we specified above.

First we will consider in abstract terms the probability that a one-bit difference in some word a produces a one-bit difference in the word d when we define $d = a + b + c$ for unknown constants b and c. One approach might be to consider this as two separate additions and to consider the intermediate word $e = a + b$ first. Since a one-bit difference in a produces a one-bit difference in e with probability 1/2 and a one-bit difference in e provides a one-bit difference in $d = e + c$ with probability 1/2 we would say that the characteristic over the two additions has probability 1/4. However it would then be misleading to use this characteristic to provide an approximation to the probability of the differential from a to d. Instead, the probability of the propagation of a one-bit difference from a to d is 1/2 since $b + c$ is a fixed value. Consequently the probability of the differential from a to d must also be 1/2.

Recall that the probability of the differential is given by the sum of the probabilities of all the characteristics that satisfy the differential. By looking at two successive additions in isolation we inadvertently restrict our attention to single-bit differences in the intermediate value e. Let α, $0 \leq \alpha \leq n - 1$, denote the position of the one bit difference in a. A one-bit difference in a will give a difference in e with Hamming weight h with probability 2^{-h}, $1 \leq h < n - \alpha$, and with probability $2^{-n+\alpha+1}$ for $h = n - \alpha$. Since this h-bit difference was caused by a one-bit difference in the previous step[1] an h-bit difference in e will be transformed to a one-bit difference in d by the addition of c with probability 1/2. Thus we get

$$p = 2^{-1}(2^{-n+\alpha} + \sum_{h=1}^{n-\alpha} 2^{-h}), \quad \text{if } \alpha < n - 1 \qquad (9)$$

$$p = 1 \text{ if } \alpha = n - 1. \qquad (10)$$

One place where this has an effect is when a mixing round follows a mashing round. Each word $R[0], \ldots, R[3]$ is modified by a `MASH` step in turn. At the first subsequent `MIX` step $R[0]$ is modified by means of addition. By looking at the two additions in isolation one under-estimates the probability of the differential.

In the analysis of RC2 we need to take account of this effect since it applies to some extent to the mixing rounds as well as during the transition between mixing

[1] In general it is not true that an h-bit difference goes to a one-bit difference with such a high probability.

and mashing rounds. Within the mixing rounds an intermediate quantity is used as input to a multiplexor function. This reduces the probability that this particular characteristic is followed by a factor of 2^{-h} for each multiplexor when the Hamming weight of the difference is h. If we denote the number of multiplexing functions between two successive additions by k then (9) can be rewritten as follows:

$$\begin{aligned} p &= \sum_{h=1}^{n-\alpha} 2^{-h} \cdot 2^{-hk} \cdot 2^{-1} + 2^{-(n-\alpha)} \cdot 2^{-(n-\alpha)k} \cdot 2^{-1} && (11) \\ &= 2^{-1} \left(\sum_{h=1}^{n-\alpha} 2^{-(k+1)h} + 2^{-(k+1)(n-\alpha)} \right) && (12) \\ &= 2^{-1} \left(\frac{1 - 2^{-(k+1)(n-\alpha)}}{1 - 2^{-(k+1)}} \frac{1}{2^{(k+1)}} \right) + 2^{-(k+1)(n-\alpha)} && (13) \\ &\approx 2^{-1} \left(\frac{1}{2^{(k+1)} - 1} \right). && (14) \end{aligned}$$

The last approximation is reasonable for smaller α ($\alpha < n-3$) but would need some correction for larger values of α. For $k = 0, 1, 2, 3$, (14) gives $p = 1/2, 1/6, 1/14, 1/30$, which should be compared with the respective probabilities of the characteristics we previously derived: $1/4, 1/8, 1/16, 1/32$. In the case of two consecutive mixing rounds we have that $k = 3$ and so the probability of a one-bit to one-bit differential across two mixing rounds is $1/30 \times 2^{-3} = 1/240$.

The effect we are using here can be extended to a series of additions whereby the intermediate values of interest have differences with a variety of Hamming weights even though the starting and ending difference have weight one. Consider three consecutive mixing rounds. Let a be a one-bit difference in the leftmost words of two inputs and let α be the position of that bit, where $0 \leq \alpha \leq n-1$. Let d be the difference in the leftmost words after three mixing rounds and suppose that h_1 and h_2 denote the Hamming weights of the leftmost words after one, respectively two, mixing rounds. Then the probability that d is a one-bit difference can be estimated as follows, where $k = 3$ and where for simplicity we have eliminated the term for $h = n - \alpha$.

$$\begin{aligned} p &\simeq \sum_{h_1=1}^{n-\alpha} \sum_{h_2=1}^{n-\alpha} 2^{-h_1} \cdot 2^{-h_1 k} \cdot 2^{-h_2} \cdot 2^{-h_2 k} \cdot 2^{-1} && (15) \\ &= 2^{-1} \left(\sum_{h_1=1}^{n-\alpha} 2^{-(k+1)h_1} \cdot \sum_{h_2=1}^{n-\alpha} 2^{-(k+1)h_2} \right) && (16) \\ &\approx 2^{-1} \left(\frac{1}{2^{(k+1)} - 1} \right) \left(\frac{1}{2^{(k+1)} - 1} \right). && (17) \end{aligned}$$

Again the final approximation requires that α is small. For $k = 3$ p is $2^{-1}(1/15)^2$. We can now estimate the probability of the differential over three mixing rounds by $2^{-1}(1/15)^2 \times 1/8 \simeq 1/3600$. This extends easily to more rounds and in general the probability of a differential over r mixing rounds is $(1/15)^{r-1} \times 1/16$. Note that the mashing rounds can be passed with probability one. Table 2 gives experimental evidence that the expressions derived are reasonable to use. The number of rounds in the table refers to the number of mixing rounds used. After five mixing rounds an additional mashing round is inserted as occurs when encrypting with RC2. The final column is derived as an average over at least five sets of experiments for each row.

Table 3 lists the probabilities of the differentials with the same input and output differences as the corresponding characteristics of Table 1, computed using the approximations of probabilities of differentials derived above.

# rounds	# pairs/test	prob.	# right pairs exp.	# right pairs obtained
3	2^{19}	$2^{-11.8}$	146	146
4	2^{22}	$2^{-15.7}$	78	79
6	2^{29}	$2^{-23.5}$	44	47
7	2^{31}	$2^{-27.4}$	12	13

Table 2: Results of experiments with differentials.

plaintext difference	difference after 15 rounds	prob.	possible values of t
$(e_t, 0, 0, 0)$	$(e_{t+15}, 0, 0, 0)$	$2^{-56.7}$	4
$(e_t, 0, 0, 0)$	$(e_{t+15}, 0, 0, 0)$	$2^{-57.7}$	1, 2, 3
$(0, e_t, 0, 0)$	$(0, e_{t+14}, 0, 0)$	$2^{-56.7}$	5
$(0, e_t, 0, 0)$	$(0, e_{t+14}, 0, 0)$	$2^{-57.7}$	1, 3
$(0, e_t, 0, 0)$	$(0, e_{t+14}, 0, 0)$	$2^{-58.7}$	0, 2, 4
$(0, 0, e_t, 0)$	$(0, 0, e_{t+13}, 0)$	$2^{-56.7}$	14
$(0, 0, e_t, 0)$	$(0, 0, e_{t+13}, 0)$	$2^{-57.7}$	7, …, 13
$(0, 0, 0, e_t)$	$(0, 0, 0, e_{t+11})$	$2^{-56.7}$	6
$(0, 0, 0, e_t)$	$(0, 0, 0, e_{t+11})$	$2^{-57.7}$	15, 0, …, 5

Table 3: 26 15-round differentials that are useful in a differential attack on RC2, which all pass the mashing rounds with probability $p \geq 1/2$.

In a differential cryptanalytic attack the attacker typically chooses a differential for $(r-1)$ rounds of an r-round block cipher. The attacker then tries to deduce key information from the last round of the cipher [3]. For RC2 the most effective attack seems to be to try and recover the bits of $K[0]$ used in the first mixing round and subsequently proceed to other rounds. In [21] a detailed analysis shows how to recover the subkey $K[0]$, however we shall not repeat the details here.

The most important factor in differential cryptanalysis is the probabilities of the differentials used in the attack. The complexity of a differential attack is approximately $c \cdot 1/p$, where p is the probability of the best differential and c a small constant.

Note that *structures* [3] can be useful in reducing the plaintext requirements for a differential attack when more than one differential is useful. With n useful differentials we can ask for a *structure* of 2^n plaintexts with specifically chosen differences. From these we derive 2^{n-1} plaintext pairs for each of the n characteristics.

In an attack on RC2 with 16 mixing rounds one would use several 15-round differentials, the probability of the best one we have detected is $2^{-56.7}$. We conjecture that no less than 2^{59} pairs are required to successfully extract the first-step key $K[0]$ and that in total a differential attack on RC2 will require at least 2^{60} chosen plaintexts.

4 Linear Cryptanalysis

The aim of a linear attack is to establish a linear relation between bits of the plaintext, bits of the corresponding ciphertext and bits of the key, a relation which holds with some probability p. Such an approximation can generally be used to provide an estimate for one bit of the key and more advanced techniques are available to extract more key information [26]. If an approximation holds with probability p then the important quantity for the cryptanalyst is the absolute value of the bias of the approximation $b = |p - 1/2|$. Typically the data required to use such an approximation is given by $c \times b^{-2}$ known plaintexts for some small constant c [26].

The `MIX` step in RC2 is

$$R[0] = R[0] + K[j] + (R[3] \,\&\, R[2]) + ({}^{\sim}R[3] \,\&\, R[1]).$$

The best linear approximations across modular additions involve the least significant bits of each quantity only, and will hold with probability one. For the multiplexor function

$$x = (R[3] \,\&\, R[2]) + ({}^{\sim}R[3] \,\&\, R[1])$$

there are linear approximations of varying qualities. The absolute value of the highest non-trivial bias is 1/4 when averaged over all inputs. It is standard notation to consider a 16-bit word x as a vector in $Z_{2^{16}}$ and to use the 16-bit quantity α to indicate the bits of x that are to be used in a linear approximation. This is most often described by means of the scalar product of two vectors. Thus the $\{0,1\}$-vector α will be used to denote the specific bits of x to be used in an approximation and $\alpha \cdot x$ is the value of these bits combined using exclusive-or. Useful linear approximations across the multiplexor are of the form

$$\begin{array}{ll} \alpha \cdot x = \alpha \cdot R[1] & \alpha \cdot x = \alpha \cdot R[1] \oplus \alpha \cdot R[3] \\ \alpha \cdot x = \alpha \cdot R[2] & \alpha \cdot x = \alpha \cdot R[2] \oplus \alpha \cdot R[3] \end{array}$$

where $\mathrm{Hwt}(\alpha) = 1$.

The following approximation to the first `MIX` step (which includes the cyclic swap of the $R[\cdot]$ words) is useful

$$e_1 \cdot (R[3]^{\mathrm{new}}) = e_0 \cdot (R[0]^{\mathrm{old}}) \quad \oplus \quad e_0 \cdot (K[j]) \oplus e_0 \cdot R[2]^{\mathrm{old}}.$$

This has a bias of absolute value 1/4. The following steps require no approximation and there appears to be no better non-trivial linear approximations for a complete mixing round. This approximation is illustrated as follows:

step	$R[0]$	$R[1]$	$R[2]$	$R[3]$	*round 1*
	e_0	$-$	e_0	$-$	*start*
1	$-$	$-$	$-$	e_1	$\lvert b\rvert = 1/4$
2	$-$	$-$	e_1	$-$	$\lvert b\rvert = 1/2$
3	$-$	e_1	$-$	$-$	$\lvert b\rvert = 1/2$
4	e_1	$-$	$-$	$-$	$\lvert b\rvert = 1/2$

In continuing this approximation into the next mixing round would require us to approximate the bit $e_1 \& R[0]$. One integer addition involves the subkey word $K[4]$ and depending on this value the bias of the approximation will vary.

The second integer addition involves the output from the multiplexor function. By the conditions given above this approximation must involve $e_1 \& R[1]$ or $e_1 \& R[2]$ and we can construct the following approximations for the second and third mixing rounds. Here we assume that the bias of the approximation across the multiplexor function is at most 1/4. Similarly, we assume that the bias of the approximation across the integer addition is at most 1/4. This occurs in approximating steps 1 and 3 and the value of $|b|$ is given for those steps individually.

step	$R[0]$	$R[1]$	$R[2]$	$R[3]$	*round 2*
	e_1	$-$	$-$	$-$	*start*
1	$-$	e_1	$-$	e_2	$\lvert b\rvert \leq 1/8$
2	e_1	$-$	e_2	$-$	$\lvert b\rvert = 1/2$
3	$-$	$e_1 \oplus e_2$	$-$	e_4	$\lvert b\rvert \leq 1/8$
4	$e_1 \oplus e_2$	$-$	e_4	$-$	$\lvert b\rvert = 1/2$

r	plaintext mask	mask after r rounds	Pul bias	Found bias	# texts
1	$(e_0, 0, e_0, 0)$	$(e_1, 0, 0, 0)$	2^{-2}	$2^{-2.0}$	2^{14}
2	$(e_0, 0, e_0, 0)$	$(e_1 \oplus e_2, 0, e_4, 0)$	2^{-6}	$2^{-5.4}$	2^{19}
2.5	$(e_0, 0, e_0, 0)$	$(e_1 \oplus e_2 \oplus e_4, 0, e_2 \oplus e_3, 0)$	2^{-9}	$2^{-8.4}$	2^{19}
3	$(e_0, 0, e_0, 0)$	$(e_1 \oplus e_3 \oplus e_4, 0, e_4 \oplus e_5 \oplus e_7, 0)$	2^{-15}	$2^{-10.3}$	2^{25}

Table 4: Biases for linear characteristics for reduced-round versions of RC2. "Pul bias" is the bias as computed with the piling-up lemma. "Found bias" is the bias found as an average in 100 using "# texts" in every test.

The typical way to measure the effectiveness of linear cryptanalysis is to appeal to the so-called *piling-up lemma* [25]. By doing this, we are lead to estimate a bias of $\leq 2^{-2} \times 2^{-3} \times 2^{-3} \times 2^{2} = 2^{-6}$ for our approximation to the first two mixing rounds of RC2. In the case of RC2, however, routine use of the piling-up lemma can lead to misleading results.

As an illustration consider Table 4. The linear approximation for one round is the one given above. In 100 tests, each with a fresh randomly chosen key, the bias of the approximation was recorded using 2^{14} texts in each test. The approximation for two rounds is the concatenation of the two characteristics reported above. As seen, using the piling-up lemma the bias is computed to 2^{-6} whereas extensive experiments indicate a bias of about $2^{-5.4}$.

Let us next consider what happens in one would try to extend the two-round approximations one additional round. This would lead to an approximation on the form:

step	$R[0]$	$R[1]$	$R[2]$	$R[3]$	*round 3*
	$e_1 \oplus e_2$	$-$	e_4	$-$	*start*
1	$-$	$e_1 \oplus e_2 \oplus e_4$	$-$	$e_2 \oplus e_3$	$\lvert b\rvert \leq 1/16$
2	$e_1 \oplus e_2 \oplus e_4$	$-$	$e_2 \oplus e_3$	$-$	$\lvert b\rvert = 1/2$
3	$-$	$e_1 \oplus e_3 \oplus e_4$	$-$	$e_4 \oplus e_5 \oplus e_7$	$\lvert b\rvert \leq 1/128$
4	$e_1 \oplus e_3 \oplus e_4$	$-$	$e_4 \oplus e_5 \oplus e_7$	$-$	$\lvert b\rvert = 1/2$

Using the piling-up lemma over all three rounds one obtaines a bias of 2^{-15}. From the last row of Table 4 it follows that the extensive experiments we performed suggest a much lower bias of about $2^{-10.3}$. To analyse the reason for this difference, the third round approximation was stopped half-way and an approximation over 2.5 rounds was tested. This approximation has a bias of 2^{-9} when computed using the piling-up lemma and about $2^{-8.4}$ in our experiments, which is shown in the second-last row in the table. Hence it appears that there is something unexpected going on the last two `MIX` steps of the third mixing round. These unexpected things probably stem from the facts that `a)` consecutive rounds in RC2 are not independent and `b)` the subkeys are not independent. This leads to complex interactions between the individual bits in the various steps of RC2 and tests provide unintuitive results.

However, even an unexpected high bias of 2^{-10} for every three rounds would not lead to a practical attack on 16-round RC2. First of all, using the bias 2^{-10} per every 3 rounds leads to a bias over 15 mixing rounds (not even considering mashing rounds) of 2^{-46}. This alone is sufficient to conclude that a linear attack is not likely over all 16 rounds of RC2. In addition, in this estimate we have ignored the mashing rounds, which will complicate things even further. And finally, it is unlikely that one can find a linear approximation over 15 rounds with a bias of 2^{-10} for every three rounds. Extending the three-round approximation used above to four rounds will involve even more bits and the bias will decrease rapidly.

5 Other cryptanalysis

In this section we consider other attacks. First of all, there are trivial attacks which apply to all block ciphers. An exhaustive key search will take 2^k operations to succeed, where k is the key size. Also, the "matching ciphertext attack" applies in ECB and CBC mode and requires about $2^{n/2}$ ciphertext blocks to succeed with good probability, where n is the block size. With $n = 64$ as in RC2, 2^{32} ciphertext blocks are required after which an attacker would be able to deduce information about the plaintext blocks.

Higher order differentials. This attack applies to ciphers which uses nonlinear components of a low algebraic degree. All components in the mixing rounds of RC2 are of (relatively) low algebraic degree but the rotations together with the modular additions is likely to make the degree after 15 mixing rounds high enough to prevent these attacks. In addition the S-box in the mashing rounds will help complicate higher order differential attacks. It is computationally infeasible to compute the exact algebraic degrees of all ciphertext bits for RC2 with more than two rounds. However, one can compute a lower bound which will give an indication of the magnitude of the exact algebraic degree. We implemented a series of tests to estimate how the algebraic degree grows in RC2. We fixed the rightmost three words of the plaintexts and ran through all other bits. In that way we can generate 2^{16} plaintexts. Then we computed the algebraic degree of all 64 ciphertext bits after $r = 1, \ldots$ rounds of encryption as a function of the bits in the leftmost 16-bit plaintext word. Our tests show that after 3 (mixing) rounds of encryption the algebraic degree of about half of the 64 ciphertext bits is the maximum 16, which is what to expect from a randomly chosen function. Similar results were obtained after three rounds, where the 16 varying bits in the plaintexts were in the second, third and fourth words. This illustrates that the algebraic degrees of the ciphertext bits grow fast in RC2, which makes higher order differential attacks very unlikely to succeed.

The slide attacks do not apply to RC2 because of the high complexity of the key-schedule.

The non-surjective attacks and the "mod n" attacks are not likely to be applicable, since the structure and components of RC2 do not seem to facilitate these attacks.

Because of the multiplexor function the integral attacks are not efficient when applied to RC2. In addition, integral attacks are usually not applicable to very many rounds.

The interpolation attacks apply to ciphers which use simple mathematical functions only. The S-box used in RC2 is generated from the number π [34]. This together with rotations, the use of boolean operations and modular additions, make the interpolation attacks very unlikely to be applicable.

The key-schedule of RC2 does not seem to allow for related-key attacks. The schedule is rather complex and involves S-box evaluations, fixed constants, rotations and both exclusive-ors and additions modulo 256. Therefore, there does not appear to be any obvious weak keys for RC2.

6 Survey of previous results on RC2

The only results we are aware of on RC2 are the results from [21], which have been included and extended above.

7 RC2 in SSL

In `SSL v2` there are two options for data encryption with RC2, one using a 40-bit key and one using a 128-bit key. One option in `SSL v3` for data encryption is RC2 with a 40-bit key. As far as we are informed, in all theses implementations RC2 is used for data encryption using the standard CBC mode.

The only potential problem with the above implementations is the use of the 40-bit keys, which give only very little security. It has been demonstrated that with special-purpose hardware a 56-bit DES keys can be found by exhaustive search in a matter of days. Moreover, a 40-bit key can be found quickly using general-purpose machines.

8 Final assessment and considerations

Of the cryptanalytic attacks considered in this report a traditional differential attack seems to be the most effective attack on RC2. The complexity of the attack is quite high and requires almost the encryptions of all possible plaintext blocks which makes the attack completely infeasible in practice.

A Block Ciphers in General

In the following we give a compressed overview of the state-of-the-art of block cipher cryptanalysis, and outline the following known attacks.

1. Exhaustive Key Search
2. Matching Ciphertext Attacks
3. Differential Cryptanalysis
4. Truncated Differential Attacks
5. Higher-order Differential Attacks
6. Linear Cryptanalysis
7. Related-key Attacks
8. Non-surjective Attacks
9. Interpolation Attacks
10. Mod-n Attacks
11. Slide Attacks
12. Integral Attacks

A.1 Exhaustive key search

This attack needs only a few known plaintext-ciphertext pairs. An attacker simply tries all keys, one by one, and checks whether the given plaintext encrypts to the given ciphertext. For a block cipher with a k-bit key and n-bit blocks the number of pairs of texts needed to determine the key uniquely is approximately $\lceil k/n \rceil$. Also, if the plaintext space is redundant, e.g., consists of English or Japanese text, the attack will work if only some ciphertext blocks is available. The number of ciphertext blocks needed depends on the redundancy of the language.

A.2 The matching ciphertext attack

The *matching ciphertext attack* is based on the fact that for block ciphers of m bits used in the modes of operations for the DES [32] after the encryption of $2^{m/2}$ blocks, equal ciphertext blocks can be expected and information is leaked about the plaintexts [6, 17, 29].

A.3 Differential cryptanalysis

The most well-known and general method of analysing conventional cryptosystems today is *differential cryptanalysis*, published by Biham and Shamir in 1990. Differential cryptanalysis is universal in the sense that it can be used against any cryptographic mapping which is constructed from iterating a fixed round function. One defines a **difference** between two bit strings, X and X' of equal length as

$$\Delta X = X \otimes (X')^{-1}, \tag{18}$$

where $\otimes$ is the group operation on the group of bit strings used to combine the key with the text input in the round function and where $(X)^{-1}$ is the inverse element of X with respect to $\otimes$. The idea behind this is, that the differences between the texts before and after the key is combined are equal, i.e., the difference is independent of the key. To see this, note that

$$(X \otimes K) \otimes (X' \otimes K)^{-1} = X \otimes K \otimes K^{-1} \otimes X'^{-1} = X \otimes (X')^{-1} = \Delta X.$$

In a differential attack one exploits that for certain input differences the distribution of output differences of the non-linear components is non-uniform.

Definition 1 *An s-round* characteristic *is a series of differences defined as an* $s+1$*-tuple* $\{\alpha_0, \alpha_1, \ldots, \alpha_s\}$*, where* $\Delta P = \alpha_0$*,* $\Delta C_i = \alpha_i$ *for* $1 \le i \le s$*.*

Here ΔP is the difference in the plaintexts and ΔC_i is the difference in the ciphertexts after i rounds of encryption. Thus, the characteristics are lists of expected differences in the intermediate ciphertexts for an encryption of a pair of plaintexts. In essence one specifies a characteristic for a number of rounds and searches for the correct key in the remaining few rounds. In some attacks it is not necessary to predict the values $\alpha_1, \ldots, \alpha_{s-1}$ in a characteristic. The pair (α_0, α_s) is called a *differential*. The complexity of a differential attack is approximately the inverse of the probability of the characteristic or differential used in the attack.

A.4 Truncated differentials

For some ciphers it is possible and advantageous to predict only the values of parts of the differences after each round of the cipher. The notion of truncated differentials was introduced by Knudsen [19]:

Definition 2 *A differential that predicts only parts of an n-bit value is called a* truncated differential. *More formally, let* (a, b) *be an i-round differential. If* a' *is a subsequence of* a *and* b' *is a subsequence of* b*, then* (a', b') *is called an i-round truncated differential.*

A truncated differential can be seen as a collection of differentials. As an example, consider an n-bit block cipher and the truncated differential (a', b), where a' specifies the least $n' < n$ significant bits of the plaintext difference and b specifies the ciphertext difference of length n. This differential is a collection of all $2^{n-n'}$ differentials (a, b), where a is any value, which truncated to the n' least significant bits is a'.

A.5 Impossible differentials

A special type of differentials are those of probability zero. The attack was first applied to the cipher DEAL [20] and later to Skipjack [2]. The main idea is to specify a differential of probability zero over some number of rounds in the attacked cipher. Then by guessing some keys in the rounds not covered by the differential one can discard a wrong value of the key if it would enable the cipher to take on the differences given in the differential.

A.6 Higher-order differentials

An sth-order differential is defined recursively as a (conventional) differential of the function specifying an $(s-1)$st order differential. In order words, an sth order differential consists of a collection of 2^s texts of certain pairwise, predetermined differences. We refer to [23, 19] for a more precise definition of higher order differentials.

In most cases one considers differences induced by the exclusive-or operation and the field of characteristic 2. The *nonlinear order* of a function $f : GF(2^n) \rightarrow GF(2^n)$ is defined as follows. Let the output bits y_j be expressed as multivariate polynomials $q_j(x) \in GF(2)[x_1, \ldots, x_n]$, where $x_1, \ldots, x_n$ are the input bits. The nonlinear order of f is then defined to be the minimum total degree of any linear combination of these polynomials. The higher order differential attacks exploit the following result.

Corollary 1 *Let $f : GF(2^n) \rightarrow GF(2^n)$ be a function of nonlinear order d. Then any dth order differential is a constant. Consequently, any $(d+1)$st order differential is zero.*

The boomerang attack [35] can be seen as a special type of a second-order differential attack. This variant applies particularly well to ciphers for which one particular (first-order) differential applies well to one half of the cipher, and where another particular (first-order) differential applies well to the other half of the cipher.

A.7 Linear cryptanalysis

Linear cryptanalysis was proposed by Matsui in 1993 [25]. A preliminary version of the attack on FEAL was described in 1992 [28]. Linear cryptanalysis [25] is a known plaintext attack in which the attacker exploits linear approximations of some bits of the plaintext, some bits of the ciphertext and some bits of the secret key. In the attack on the DES (or on DES-like iterated ciphers) the linear approximations are obtained by combining approximations for each round under the assumption of independent round keys. The attacker hopes in this way to find an expression

$$(P \cdot \alpha) \oplus (C \cdot \beta) = (K \cdot \gamma) \tag{19}$$

which holds with probability $p_L \neq \frac{1}{2}$ over all keys [25], such that $|p_L - \frac{1}{2}|$, called the bias, is maximal. In (19) $P, C, \alpha, \beta, \gamma$ are m-bit strings and '$\cdot$' denotes the dot product. The bit strings α, β, γ are called *masks.*

Definition 3 *An s-round* linear characteristic *is a series of masks defined as an $(s+1)$-tuple $\{\alpha_0, \alpha_1, \ldots, \alpha_s\}$, where α_0 is the mask of the plaintexts and α_i is the mask of the ciphertexts after i rounds of encryption for $1 \leq i \leq s$.*

As for differential cryptanalysis one specifies a linear characteristics for a number of rounds and searches for the keys in the remaining rounds, we refer to [25] for more details. A linear attack needs approximately about b^{-2} known plaintexts to succeed, where b is the bias of the linear characteristic used.

Also, the concepts of linear hulls, the analogue to differentials as opposed to characteristics in differentials cryptanalysis, has been defined in [31].

Finally, in [27] it has been shown that if one defines the quantity $q = (2p - 1)^2$ where p is the probability of a linear characteristic or hull, then when combining several linear characteristics one can multiply their q values to get the q-value of the combination. Sometimes the q values are referred to as the "linear probability", which is somewhat misleading, but nevertheless seems to be widely used.

A.8 Mod n cryptanalysis

In [15] a generalisation of the linear attacks is considered. This attack is applicable to ciphers for which some words (in some intermediate ciphertext) are biased modulo n, where n typically is a small integer. It has been shown that ciphers which uses only bitwise rotations and additions modulo 2^{32} are vulnerable to these kinds of attacks.

A.9 Related-key attacks

There are several variants of this attack depending on how powerful the attacker is assumed to be.

1. Attacker gets encryptions under one key.
2. Attacker gets encryptions under several keys.
   (a) Known relation between keys.
   (b) Chosen relation between keys.

Knudsen used the methods of 1 by giving a chosen plaintext attack of the first kind on LOKI'91 [16], reducing an exhaustive key search by almost a factor of four. The concept "related-key attack" was introduced by Biham [1], who also introduced the attack scenarios of 2, where the encryptions under several keys are requested. Knudsen later described a related key attack on SAFER K [18] and Kelsey, Schneier, and Wagner [14] applied the related key attacks to a wide range of block ciphers. It may be argued that the attacks with a chosen relation between the keys are unrealistic. The attacker need to get encryptions under several keys, in some attacks even with chosen plaintexts. However there exist realistic settings, in which an attacker may succeed to obtain such encryptions. Also, there exists quite efficient methods to preclude the related key attacks [14, 10].

A.10 Interpolation attack

In [12] Jakobsen and Knudsen introduced the interpolation attack on block ciphers. The attack is based on the following well-known formula. Let R be a field. Given $2n$ elements $x_1, \dots, x_n, y_1, \dots, y_n \in R$, where the x_is are distinct. Define

$$f(x) = \sum_{i=1}^{n} y_i \prod_{1 \le j \le n, j \ne i} \frac{x - x_j}{x_i - x_j}. \tag{20}$$

$f(x)$ is the only polynomial over R of degree at most $n - 1$ such that $f(x_i) = y_i$ for $i = 1, \dots, n$. Equation (20) is known as the *Lagrange interpolation formula* (see e.g.,[5, page 185]). In the *interpolation attack* an attacker constructs polynomials using pairs of plaintexts and ciphertexts. This is particularly easy if the components in the cipher can be expressed as easily described mathematical functions. The idea of the attack is, that if the constructed polynomials have a small degree, only few plaintexts and their corresponding ciphertexts are necessary to solve for the (key-dependent) coefficients of the polynomial, e.g., using Lagrange's interpolation. To recover key bits one expresses the ciphertext before the last round as a polynomial of the plaintext.

A.11 Non-surjective attack

In [33] Rijmen-Preneel-De Win described the non-surjective attack on iterated ciphers. It is applicable to Feistel ciphers where the round function is not surjective and therefore statistical attacks become possible. In a Feistel cipher one can compute the exclusive-or of all outputs of the round functions from the plaintexts and the corresponding ciphertexts. Thus, if the round functions are not surjective this gives information about intermediate values in the encryptions, which can be used to get information about the secret keys.

A.12 Slide attacks

In [4] the "slide attacks" were introduced, based on earlier work in [1, 16]. In particular it was shown that iterated ciphers with identical round functions, that is, equal structures plus equal subkeys in the rounds, are susceptible to slide attacks. Let $F_r \circ F_{r-1} \circ \cdots \circ F_1$ denote an r-round iterated cipher, where all F_is are identical. The attacker tries to find pairs of plaintext P, P^* and their corresponding ciphertexts C, C^*, such that $F_1(P) = P^*$ and $F_r(C) = C^*$. Subsequently, an attacker has twice both the inputs and outputs of one round of the cipher. If the round function is simple enough, this can lead to very efficient attacks. To find such pairs of texts, one can in the worst case apply the birthday paradox, such that one such pair is expected from a collection of $2^{n/2}$ texts, where n is the block size.

A.13 Integral Attacks

These attacks are sometimes referred to as the "Square attack", since it was first applied to the block cipher Square [7, 8]. The attack on Square slightly modified also applies to the block ciphers Crypton and Rijndael [9].

In [22] these attacks are generalised under the name of "integral cryptanalysis". In differential attacks one considers differences of texts, in integral cryptanalysis one considers sums of texts. In ciphers where all nonlinear functions are bijective, it is sometimes possible to predict a sum of texts, even in the cases where differential attacks are not applicable. The main observations are that in a collection of texts which in a particular word take all values exactly equally many times, the value of the words after a bijective function also take all values exactly equally many times. Also, assume that s words have this property and that in the cipher a linear combination of the s words are computed (with respect to the group operation considered). Then it is possible to determine also the sum of all linear combinations in a collection of texts. This attack is still today the best attack reported on Rijndael which has been the selected for the Advanced Encryption Standard.

References

[1] E. Biham. New types of cryptanalytic attacks using related keys. In T. Helleseth, editor, *Advances in Cryptology: EUROCRYPT'93, LNCS 765*, pages 398–409. Springer Verlag, 1993.

[2] E. Biham, A. Biryukov, and A. Shamir. "Impossible" cryptanalysis. Presented at the rump session of CRYPTO'98.

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

[4] A. Biryukov and D. Wagner. Slide attacks. In L. R. Knudsen, editor, *Fast Software Encryption, Sixth International Workshop, Rome, Italy, March 1999, LNCS 1636*, pages 245–259. Springer Verlag, 1999.

[5] P.M. Cohn. *Algebra, Volume 1.* John Wiley & Sons, 1982.

[6] D. Coppersmith, D.B. Johnson, and S.M. Matyas. Triple DES cipher block chaining with output feedback masking. Technical Report RC 20591, IBM, October 1996. Presented at the rump session of CRYPTO'96.

[7] J. Daemen, L. Knudsen, and V. Rijmen. The block cipher Square. In E. Biham, editor, *Fast Software Encryption, Fourth International Workshop, Haifa, Israel, January 1997, LNCS 1267*, pages 149–165. Springer Verlag, 1997.

[8] J. Daemen, L. Knudsen, and V. Rijmen. The block cipher Square. Dr. Dobbs Journal, October 1997.

[9] J. Daemen and V .Rijmen. AES proposal: Rijndael. Submitted as an AES Candidate Algorithm. Description available from NIST, see `http://www.nist.gov/aes`.

[10] I.B. Damgård and L.R. Knudsen. Two-key triple encryption. *The Journal of Cryptology*, 11(3):209–218, 1998.

[11] S. Dusse, P. Hoffman, B. Ramsdell, L. Lundblade, and L. Repka. S/MIME Message Specification. September 23, 1997. Available from http://www.imc.org/draft-dusse-smime-msg.

[12] T. Jakobsen and L. Knudsen. The interpolation attack on block ciphers. In E. Biham, editor, *Fast Software Encryption, Fourth International Workshop, Haifa, Israel, January 1997, LNCS 1267*, pages 28–40. Springer Verlag, 1997.

[13] B. Kaliski and Y.L. Yin. On differential and linear cryptanalysis of the RC5 encryption algorithm. In D. Coppersmith, editor, *Advances in Cryptology: CRYPTO'95, LNCS 963*, pages 171–184. Springer Verlag, 1995.

[14] J. Kelsey, B. Schneier, and D. Wagner. Key-schedule cryptanalysis of IDEA, G-DES, GOST, SAFER, and triple-DES. In Neal Koblitz, editor, *Advances in Cryptology: CRYPTO'96, LNCS 1109*, pages 237–251. Springer Verlag, 1996.

[15] J. Kelsey, B. Schneier, and D. Wagner. Mod n cryptanalysis, with applications against RC5P and M6. In L. Knudsen, editor, *Fast Software Encryption, Sixth International Workshop, Rome, Italy, March 1999, LNCS 1636*, pages 139–155. Springer Verlag, 1999.

[16] L.R. Knudsen. Cryptanalysis of LOKI'91. In J. Seberry and Y. Zheng, editors, *Advances in Cryptology, AusCrypt 92, LNCS 718*, pages 196–208. Springer Verlag, 1993.

[17] L.R. Knudsen. *Block Ciphers – Analysis, Design and Applications.* PhD thesis, Aarhus University, Denmark, 1994.

[18] L.R. Knudsen. A key-schedule weakness in SAFER K-64. In Don Coppersmith, editor, *Advances in Cryptology - CRYPTO'95, LNCS 963*, pages 274–286. Springer Verlag, 1995.

[19] L.R. Knudsen. Truncated and higher order differentials. In B. Preneel, editor, *Fast Software Encryption - Second International Workshop, Leuven, Belgium, LNCS 1008*, pages 196–211. Springer Verlag, 1995.

[20] L.R. Knudsen. DEAL - a 128-bit block cipher. Technical Report 151, Department of Informatics,University of Bergen, Norway, February 1998. Submitted as an AES candidate by Richard Outerbridge.

[21] L.R. Knudsen, V. Rijmen, R.L. Rivest, and M.P.J. Robshaw. On the design and security of RC2. In S. Vaudenay, editor, *Fast Software Encryption, Fifth International Workshop, FSE'98, Paris, France, March 1998, LNCS 1372*, pages 206–221. Springer Verlag, 1998.

[22] L.R. Knudsen, D. Wagner. Integral cryptanalysis. To be presented at *Fast Software Encryption*, Leuven, Belgium, February 2002. To appear in proceedings from Springer Verlag.

[23] X. Lai. Higher order derivatives and differential cryptanalysis. In R. Blahut, editor, *Communication and Cryptography, Two Sides of One Tapestry.* Kluwer Academic Publishers, 1994. ISBN 0-7923-9469-0.

[24] X. Lai, J.L. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. In D.W. Davies, editor, *Advances in Cryptology - EUROCRYPT'91, LNCS 547*, pages 17–38. Springer Verlag, 1992.

[25] M. Matsui. Linear cryptanalysis method for DES cipher. In T. Helleseth, editor, *Advances in Cryptology - EUROCRYPT'93, LNCS 765*, pages 386–397. Springer Verlag, 1993.

[26] M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In Y.G. Desmedt, editor, *Advances in Cryptology - CRYPTO'94, LNCS 839*, pages 1–11. Springer Verlag, 1994.

[27] M. Matsui. New structure of block ciphers with provable security against differential and linear cryptanalysis. In D. Gollman, editor, *Fast Software Encryption, Third International Workshop, Cambridge, UK, February 1996, LNCS 1039*, pages 205–218. Springer Verlag, 1996.

[28] M. Matsui and A. Yamagishi. A new method for known plaintext attack of FEAL cipher. In R. Rueppel, editor, *Advances in Cryptology - EUROCRYPT'92, LNCS 658*, pages 81–91. Springer Verlag, 1992.

[29] U.M. Maurer. New approaches to the design of self-synchronizing stream ciphers. In D.W. Davies, editor, *Advances in Cryptology - EUROCRYPT'91, LNCS 547*, pages 458–471. Springer Verlag, 1991.

[30] National Institute of Standards and Technology. Advanced encryption algorithm (AES) development effort. `http://www.nist.gov/aes`.

[31] K. Nyberg. Linear approximations of block ciphers. In A. De Santis, editor, *Advances in Cryptology - EUROCRYPT'94, LNCS 950*, pages 439–444. Springer Verlag, 1995.

[32] National Bureau of Standards. DES modes of operation. Federal Information Processing Standard (FIPS), Publication 81, National Bureau of Standards, U.S. Department of Commerce, Washington D.C., December 1980.

[33] V. Rijmen, B. Preneel, and E. De Win. On weaknesses of non-surjective round functions. *Designs, Codes, and Cryptography*, 12(3):253–266, 1997.

[34] R.L. Rivest. A Description of the RC2$^{\mathrm{TM}}$ Encryption Algorithm. Available from `http://www.ietf.org/rfc/rfc2268.txt`.

[35] D. Wagner. The boomerang attack. In L. R. Knudsen, editor, *Fast Software Encryption, Sixth International Workshop, Rome, Italy, March 1999, LNCS 1636*, pages 156–170. Springer Verlag, 1999.