Analysis of SC2000

January 12, 2001

Executive Summary

This report presents the results of a limited time evaluation of the block cipher SC2000. The evaluation consisted of theoretical derivations and practical experimentations.

No flaws nor weaknesses have been identified in the design which could lead to cryptanalytic attacks with respect to the state-of-the-art. However, the design is complicated and uses components which do not facilitate for easy analysis. In particular, it seems hard to evaluate the exact level of security against the differential and linear attacks.

The designers' own analysis is, although comprehensive, not exhaustive. Nevertheless, it is felt that a cryptanalytic attack on SC2000 which can find the key faster than by an exhaustive search, is either of a high complexity or it would require a new attack which is unknown today.

Finally we mention that this report is the result of a limited time of review, and the analysis was performed without access to computer code implementing the block cipher. A concentrated, longer analysis might reveal properties of SC2000 which we were not able to detect.

Contents

1 Structural features and characteristics

SC2000 is an iterated block cipher with 128-bit blocks and allows for three different key sizes to be compliant with the AES [31]. The design consists of elements also used in the block cipher Serpent [3] and there are features reminiscent of the block cipher Twofish [33]. Unlike most other iterated block ciphers SC2000 uses two different round structures. One is the Feistel structure, the other is the SP-structure. Both these structures are classical, but it is unusual to see both used in the same block cipher. Although this may complicate all known attacks, it unfortunately also complicates the designers' analysis and it is hard to be convinced about the strength of the algorithm.

The individual components are the I-function, the B-function and the R-function. One round of SC2000 is defined as the concatenation of one I-function, one B-function, then again one I-function and two R-functions. For 128-bit keys, six such rounds are specified together with an output transformation which consists of one I-function, one B-function and finally one I-function. This special structure is used so decryption is similar to encryption. For 192-bit and 256-bit keys seven rounds are specified together with the output transformation.

It is unclear to us why the designers chose to use a mix of two different structures and there seems to be no clear arguments why this is advantageous compared to staying with one of the two structures. On the other hand, after this limited time review of SC2000 there seems to be no flaws or short-cuts in this design either.

The key-schedule is very complicated and from the designers tests on speed one evalutation of the key-schedule takes the time of more than one encryption or decryption. In scenarios where the key changes often this is an undesirable property. What speaks in favor of the key schedule design is that it seems to effectively thwart the related-key attacks and there are no (easily identified) weak keys.

2 Differential cryptanalysis

In the following we evaluate the individual components of SC2000 with respect to differential cryptanalysis. A difference of two bit-strings of equal length is defined via the exclusive-or operation.

The I-function is the exclusive-or of key material, which by itself is trivially weak in differential cryptanalysis. Therefore in the following when we consider a B-function it is understood that this is preceded by an I-function, and when we consider a pair of R-functions it is understood that these are preceded by an I-function.

The B-function

The B-function evaluates 32 instances of the S-box, S_4, which can be done in parallel using the bit-slice techniques. Each S-box is a bijective function which takes a 4-bit input and produces a 4-bit output. The highest probability of a one-round characteristic is $\frac{1}{4}$ and there are many of these. That is, consider a pair of 128-bit inputs to the B-function, such that they differ in only one bit. Then due to the nature of the B-function, the outputs will differ in at least one bit and in at most four bits. This means that there are (many) truncated differentials of probability one through one B-function.

Iterating the B-function (together with an I-function) will result in a very weak cryptosystem. A pair of 128-bit inputs different in only one bit will be different in at most four bits after any number of rounds and there are truncated differentials of probability one over any number of rounds.

For S_4 it holds that 4-bit inputs of difference $(x, y, 0, 0)$ where x and y are not both zero, never result in a difference $(v, w, 0, 0)$ for any values of v and w. And similarly, 4-bit inputs of difference $(0, 0, x, y)$ where x and y are not both zero, never result in a difference $(0, 0, v, w)$ for any values of v and w. However, there are combinations possible where 4-bit inputs different in one bit result in outputs different in one bit. Thus, there is no avalanche in B itself.

The R-function

The R-function has a one-round Feistel structure, such that in any two consecutive R-rounds the halves are swapped after the first R-round but not after the second. We shall call this the RR-construction.

First we study the different subcomponents of the F-function. First the 64-bit input is split into two halves of 32 bits each. Each half is evaluated through six S-boxes, two of 6 bits and four of 5 bits. Each of the two 32-bit outputs of the S-boxes is input to the M-function. The outputs of the M-function are then input to the L-function, the output of which forms the output of the F-function.

The S-boxes. S_5 and S_6 are both constructed from a power polynomial.

S_5 is constructed from the power polynomial x^7 in $GF(2^5)$ together with an affine mapping. The maximum probability of a non trivial characteristic through S_5 is 2^{-4}, which was confirmed by computer experiments.

S_6 is constructed from the power polynomial $x^{62} = x^{-1}$ in $GF(2^6)$ together with an affine mapping. The maximum probability of a non trivial characteristic through S_6 is also 2^{-4} [28].

For both S-boxes the design principles do not exclude the existence of characteristics for which inputs different in only one bit result in outputs different in only one bit.

The M-function. The M-function has the property for differential cryptanalysis that if there are s active S-boxes in the inputs, then there are at least $\max(1, 6 - s)$ active S-boxes in the outputs [2]. For all values of s there are examples of characteristics meeting the bounds of the active S-boxes in the outputs [2].

The L-function. The L-function takes as input two 32-bit values a and b and returns two 32-bit values c and d, such that

$$c = a \text{ AND mask} \oplus b \quad (1)$$

$$d = b \text{ AND } \overline{\text{mask}} \oplus a, \quad (2)$$

where “mask” is a predetermined 32-bit constant of Hamming weight 16. Here we list some properties of the L-function.

if $(a \text{ AND mask}) = 0$ and $b = 0$ then $(c, d) = (0, a)$,
if $(a \text{ AND mask}) = a$ and $b = 0$ then $\overline{(c, d)} = (a, a)$,
if $(a \text{ AND mask}) = 0$ and $(b \text{ AND } \overline{\text{mask}}) = 0$ then $(c, d) = (b, a)$.

Reversing the role of a and b in the first two rules, will give similar results. Since the L-function is linear with respect to the exclusive-or operation, the above properties will hold also, when a, b, c, d represent values of exclusive-or differences.

Consider the RR-construction and a pair of 128-bit inputs to the first R-function, such that they differ in only one bit in the left 64-bit halves. Then due to the nature of the R-function, the right 64-bit halves of the outputs will differ in only one bit and the left halves will be equal. Consider now the second R-function. Here the inputs differ only in one bit in the right 64-bit halves. Then the left halves of the outputs of the R-function differ also only in one bit. Also, the right 64-bit halves of the outputs will differ in at least five bits. To see this, consider two inputs to the F-function which differ only in one bit. Then it is possible that the outputs of the S-boxes differ in only one bit. According to the above mentioned properties of M, it is possible that the outputs of M differ in only five bits. Thus in one of the two inputs to the L-function the texts differ in five bits and in the other input the texts are equal. Then if the condition of the above first property of the L-function is satisfied the outputs of the L-function differ in only five bits. In total, for the RR-construction according to the design principles of SC2000 there could exist characteristics where inputs different in only one bit result in outputs different in only six bits.

Iterating the RR-construction without considering the B-function does not yield a strong cryptosystem, mainly because the halves are not swapped after the second R-function. Let us shortly analyse the cryptosystem resulting from an iteration of the RR-construction but where the halves after swapped after each R-function. This cryptosystem has a classical Feistel structure, let us denote this by RR*. Let us consider 12 rounds of the cryptosystem RR*. This corresponds to SC2000 but where the B-functions are replaced by a swapping of the 64-bit halves. Thus, this cryptosystem is faster than SC2000. For this system it is possible to make some relatively easy and crude estimations of the probabilities of characteristics. In a three-round iterative characteristic the total number of active S-boxes will be at least six. In [2] the results of a search for such characteristics or differentials showed a minimum of seven active S-boxes. In a four-round iterative characteristic the total number of active S-boxes is at least seven. However, when concatentated with itself the total number of active S-boxes is at least eleven. In total, this is eighteen active S-boxes for eight rounds. Thus a very conservative bound for the number of active S-boxes for this system is two per round. For a 12-round cryptosystem this means that a characteristic will have a probability of at most $(2^{-8})^{12} = 2^{-96}$. We conjecture that in reality this number will be much lower.

The BRR-construction

Let us next consider the construction used in SC2000. By the BRR-construction we shall mean, the application of one I-function, one B-function, then one I-function and finally one RR-construction.

From the above results on the B-function and for the RR-construction, it follows that there are characteristics for the BRR-construction for which inputs different in one bit result in outputs different in only six bits. Thus, the avalanche effect in the RR-construction is perhaps not as big as one might expect, taking into consideration the computational complexity of the functions involved. Iterating the RR-construction (together with a swap of halves after the second R-function) results in a Feistel cipher for which crude bounds of the probabilities of characteristics are relatively easy to establish. For the BRR-construction this does not seem to be the case. One needs to consider how the B-function and the RR-construction interact. The design principle of the B-function above, the fact that certain one-round characteristics have probability zero, is not immediately clear. However, by a closer look at differential attacks the reason becomes clearer. Since the halves are not swapped after the second R-function, the role of the B-function seems to be to ensure that there is an effect “similar” to a swapping of the halves. In the cryptosystem RR*, defined above, a two-round iterative characteristic is not possible, since the F-function is bijective. With the introduction of an (arbitrary) B-function this is no longer guaranteed. In fact, as illustrated by the authors [2] the special properties of the B-function do not prevent the existence of characteristics which repeat themselves in the RR-construction after two such rounds. Let us illustrate this. Denote by (x_0, x_1, x_2, x_3) the difference in two 128-bit texts, split into four 32-bit quantities. We shall write

$$(x_0, x_1, x_2, x_3) \overset{G}{\to} (y_0, y_1, y_2, y_3)$$

if texts of differences (x_0, x_1, x_2, x_3) can result texts of differences (y_0, y_1, y_2, y_3) after one application of a function G. Then the following type of characteristic is possible

$$\begin{aligned} (c, d, a, b) &\overset{B}{\to} (a, b, 0, 0) \\ (a, b, 0, 0) &\overset{R}{\to} (0, 0, a, b) \\ (0, 0, a, b) &\overset{R}{\to} (c, d, a, b), \end{aligned}$$

where a and b are not both zero. Here the probability in the first round of R is one. This characteristic could be iterated to any number of rounds. However it has been argued that characteristics of this type have very low probabilities when iterated to the whole cipher [2].

The best analytical strategy in differential cryptanalysis that we have found is the following. Consider the following characteristic

$$\begin{aligned}(a_1, b_1, c_1, d_1) &\overset{B}{\rightarrow} (a_2, b_2, c_2, d_2)\\ (a_2, b_2, c_2, d_2) &\overset{R}{\rightarrow} (a_3, b_3, c_3, d_3)\\ (a_3, b_3, c_3, d_3) &\overset{R}{\rightarrow} (a_4, b_4, c_4, d_4).\end{aligned}$$

The idea is to find values of (a_i, b_i, c_i, d_i) such that it holds that

$$h_i = a_i \text{ OR } b_i \text{ OR } c_i \text{ OR } d_i$$

for $i = 1, 4$, have lowest possible Hamming weights, and such that (c_2, d_2) and (c_3, d_3) are of a form minimising the number of active S-boxes in the F-function. Let us denote by $H(x)$ the Hamming weight of x. The values $H(h_1)$ and $H(h_4)$ give the exact number of active S-boxes in the B-function at the input to the round and the number of active S-boxes in the B-function starting in the next round. One could then analytically examine the situations occurring when demanding that $H(h_1)$ and $H(h_4)$ have certain low values. Let us start by considering the simple case where $H(h_1) = H(h_4) = 1$. Then there are two situations to examine. The first case: the inputs to the first F-function could be equal, in which case the outputs are equal. But then the inputs to the second F-function will have at least one and at most two active S-boxes, in the latter case the active S-boxes will be in different halves of the F-function. This means that in the outputs of the M-function in the second F-function there will be five or ten active S-boxes. In both cases, the L-function ensures that the outputs of the F-function have five active S-boxes. But then $H(h_4)$ cannot be 1, so this case is never reached. The second case: the inputs to the first F-function are different in one or two bits. This means that the inputs to the first F-function will have at least one and at most two active S-boxes, in the latter case the active S-boxes will be in different halves of the F-function. Similarly to before there will be at least 5 active S-boxes in the outputs of the F-function. But now the analysis becomes complicated. In the outputs to the F-function to the second round, there can be from 3 to 7 active S-boxes. Some of the cases cannot occur since it is required that $H(h_4) = 1$. The remaining cases have to take the S-boxes S_5 and S_6, the M-function and the L-function into account. We think it is possible to continue this analysis and it is likely that it is not possible to achieve $H(h_1) = H(h_4) = 1$. Even if it would be possible, there is no guarantee that the characteristic would be iterative, so one would have to do a similar analysis for a second round. Clearly, one should continue the analysis for other values of $H(h_1)$ and $H(h_4)$, but for increasing values the analysis becomes more and more complex. In [2] the authors have conducted a semi-search for so-called truncated characteristics, where only the number of active S-boxes are counted in a conversative way. The best characteristic found by the designers of SC2000 had $H(h_1) = H(h_4) = 9$.

A crude estimate would be three active S-boxes per B-round and three active S-boxes per R-round. For this to happen one needs to put many constraints on all S-boxes plus the M-function and the L-function. Let us consider this case in some more details. Clearly, three active S-boxes through the B-function is possible. The question is how does the characteristic perform through the R-functions. Let us assume that each of the two halves of the inputs to F differ in three bits, such that each pair of inputs to an S-box are either equal or differ only in one bit. Then it is possible that this is also the case after the applications of the S-boxes. The M-function will mix the bits, but according to the above mentioned properties of the M-function it is possible that for each of the two halves of outputs there will be only three active S-boxes. Assume that this "activeness" stems from a difference in only three bits in certain positions. If it is further assumed that the positions where the one-bit differences occur are identical in the two halves, then after the L-function for each of the two halves of outputs the texts could differ in only three bits. Thus, this characteristic is a possible, although unlikely, way to go through the BRR-construction. It requires that special entries in the S-boxes have nonzero probabilities and it requires some properties of the M-function which are unlikely to exist. However, by disregarding these constraints we can get an estimate of the probability of such a characteristic and use this as a lower bound. With three active S-boxes in both S_4, S_5, and in S_6, the probability of a characteristic for the BRR-construction is already bounded by $(2^{-4})^6(2^{-2})^3 = 2^{-30}$, which iterated to five BRR-constructions, i.e., 15 rounds of SC2000, yields a probability of 2^{-150}. First of all, this probability is computed using the best possible combinations through the three S-boxes in every estimation, plus disregarding big problems caused by the M-function and the L-function. If they were (or could be) included, they would decrease the probability even further.

Differentials and truncated differentials

The above analysis considered characteristics, see Appendix. Even stronger tools are differentials and truncated differentials. In the following we consider truncated differentials, but often refer to them as simply, differentials. It is possible that there exist several characteristics which can be combined into a differential. There have been block cipher cases in the past, where the differentials have a much higher probability than for corresponding characteristics. In the following we discuss if it is possible to boost the probability of the above characteristics for SC2000 by considering differentials. First of all, the outline of a characteristic and a differential are generally the same. Thus, let us consider the differential on a form similar to the above for characteristics, that is, the following form

$$\begin{array}{ccc} (a_1, b_1, c_1, d_1) & \overset{R}{\to} & (a_2, b_2, c_2, d_2) \\ (a_2, b_2, c_2, d_2) & \overset{R}{\to} & (a_3, b_3, c_3, d_3) \\ (a_3, b_3, c_3, d_3) & \overset{B}{\to} & (a_4, b_4, c_4, d_4), \end{array}$$

Since the only conditions for the differentials are some restrictions on the Hamming weights of the involved differences, there could be several possible characteristics all resulting in the same differential. There are three active S-boxes in the B-function, so the Hamming weight of the inputs to the B-function is between three and twelve. The only thing required about the outputs are that the inputs to the F-function in the first R-function have three active S-boxes in each half. This is a restriction of only six of the possible twelve bits. Thus, there could be $2^6 = 64$ possible "good" output differences. Some of these will not be possible, due to certain constraints put on the S-box S_4 in the B-function. In the first R-function, if the outputs are as desired then they are likely to have a particular and special form, such that when exclusive-ored to the left halves of the input to the R-function, they form a "good" input difference to the second F-function. Thus, for any of the possible 64 good output differences from the B-function, there seems to be at most one good after the first F-function. After the second F-function, there could be more possibilities again, since it is only required that for the subsequent B-function there will be three active S-boxes. It is very difficult to estimate the probabilities of this truncated differential, since it has to take into account all specific details of all functions and S-boxes involved. However, an optimistic prediction could be that there are totally 32 characteristics which together form a truncated differential. The upper bound on the probability of each these characteristics is 2^{-30}, cf. above, but they are likely to be much less. Thus, a conservative estimate is that this truncated differential iterated to 15 rounds of SC2000 will have a probability of **at most** 2^{-125}. Note that the probabilities of truncated differentials must, in general, be higher than 2^{-128} to make any sense in an attack. The reason for this is, that the differential only specifies the exact values in a subset of all 128 bits, whereas the remaining bits can take any values. As an example, consider the above differential for 15 rounds of SC2000. There are 12 bits in the differential which are not predicted. The remaining bits are expected to have zero values in the differential. Thus for a randomly chosen permutation there is such a differential with a probability of 2^{-116}. Thus it seems that truncated differentials as considered in this paper are of no threat for the security of SC2000.

Finally it should be stressed that the above differential analysis is by no means exhaustive. It is also felt that the structure of SC2000 is so complex that determining the exact security level of the algorithm with respect to differential cryptanalysis is a near impossible task.

3 Linear cryptanalysis

In the following we evaluate the individual components of SC2000 with respect to linear crypanalysis in a manner similar to that in the previous section for differential cryptanalysis. In the following by "linear probability" we shall mean the quantity $(2p-1)^2$, where p is the probability of the involved linear approximation, see also the Appendix.

The B-function

The highest linear probability of a non trivial one-round linear characteristic through the S-box S_4 in the B-function is $\frac{1}{4}$, which was confirmed by computer experiments. There exist linear relations with one-bit input masks and one-bit output masks.

The R-function

The S-boxes. The highest linear probability of non-trivial one-round linear characteristics through the S-boxes S_5 and S_6 in the R-function are in both cases $\frac{1}{16}$. The first case was confirmed by computer experiments, the second case follows from [28]. For both S-boxes it is possible that a linear relation exists with a one-bit input mask and a one-bit output mask.

The M-function. For linear attacks the M-function has a property similar as for differential cryptanalysis that if there are s active S-boxes in the inputs, then there are at least $\max(1, 6-s)$ active S-boxes in the outputs [2].

The L-function. Consider the properties of the L-function listed above for differential cryptanalysis. Since the L-function is linear the above properties will hold also, when a, b, c, d represent values of masks to be used in linear cryptanalysis.

The BRR-construction

In [5] it was shown that differential and linear characteristics have a dual property. As can be seen the above properties of SC2000 with respect to linear attacks are very similar to the properties for a differential attacks. Thus, the constructions of differential characteristics as above can be translated into the linear characteristics.

A crude estimate is therefore again that for the BRR-construction there will be at least three active S-boxes for each B-function and each R-function. Thus the probability of a linear characteristic for the BRR-construction is bounded by $(2^{-4})^6(2^{-2})^3 = 2^{-30}$, which iterated to five BRR-constructions, yields a probability of 2^{-150}. Again, this is so low that there is no hope that with these characteristics, if they exist at all, a linear attack will be possible. Similar to the above considerations about differentials versus characteristics in differential cryptanalysis, one can do estimates on the effect of considering linear hulls. The conclusion drawn are the same as for differential cryptanalysis.

4 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, but requires about $2^{n/2}$ ciphertext blocks to succeed with good probability, where n is the block size. With $n = 128$ as in SC2000, 2^{64} 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. SC2000 uses S-boxes of a high nonlinear order and together with the relatively high number of rounds, the probability that a higher order differential attack could be applicable is very small.

The slide attacks, the non-surjective attacks and the "mod n" attacks do not seem applicable to SC2000 .

The integral attacks can be applied to SC2000 to some extent, but it is estimated that such an approach will be much less effective than one based on differentials.

The interpolation attacks apply to ciphers which use simple mathematical functions only. SC2000 uses mathematical functions in the S-boxes, however the different natures of the B-function and of the R-function together with the affine mappings in the S-boxes have a good effect in thwarting the interpolation attacks.

The key-schedule of SC2000 uses components of the encryption module to generate the subkeys in a rather complex manner. Therefore, there is no reason to suspect that the related-key attacks are applicable nor that there are any particularly weak keys.

5 Survey of previous results

The only previous results on SC2000 that we are aware of are those of the designers themselves [2].

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 [30] after the encryption of $2^{m/2}$ blocks, equal ciphertext blocks can be expected and information is leaked about the plaintexts [9, 18, 27].

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

$$\Delta X = X \otimes (X')^{-1}, \tag{3}$$

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 $1 \leq i \leq s$.*

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 [20]:

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 [21] and later to Skipjack [6]. 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, 20] 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 $f : GF(2^n) \to GF(2^n)$ 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 $f : GF(2^n) \to GF(2^n)$ 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 [34] 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 [24]. A preliminary version of the attack on FEAL was described in 1992 [26]. Linear cryptanalysis [24] 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

$$(P \cdot \alpha) \oplus (C \cdot \beta) = (K \cdot \gamma) \tag{4}$$

which holds with probability $p_L \neq \frac{1}{2}$ over all keys [24], such that $|p_L - \frac{1}{2}|$, called the bias, is maximal. In (4) $P, C, \alpha, \beta, \gamma$ 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 [24] 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 [29].

Finally, in [25] 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 [16] 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 [17], reducing an exhaustive key search by almost a factor of four. The concept "related-key attack" was introduced by Biham [4], 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 [19] and Kelsey, Schneier, and Wagner [15] 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 [15, 13].

A.10 Interpolation attack

In [14] 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 $x_1, \ldots, x_n$, $y_1, \ldots, y_n \in R$, where the x_is are distinct. Define

$$f(x) = \sum_{i=1}^{n} y_i \prod_{1 \leq j \leq n, j \neq i} \frac{x - x_j}{x_i - x_j}. \tag{5}$$

$f(x)$ is the only polynomial over R of degree at most $n-1$ such that $f(x_i) = y_i$ for $i = 1, \ldots, n$. Equation (5) is known as the *Lagrange interpolation formula* (see e.g.,[8, 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 [32] 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 [7] the "slide attacks" were introduced, based on earlier work in [4, 17]. 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 [11, 10]. The attack on Square slightly modified also applies to the block ciphers Crypton and Rijndael [12].

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] Shimoyama, Yanami, Yokohama, Takenaka, Itoh, Yajima, Torii, Tanaka. The Block Cipher SC2000. Cryptographic Techniques Specification. Fujitsu Laboratories LTD and Science University of Tokyo. July 11, 2000.

[2] Shimoyama, Yanami, Yokohama, Takenaka, Itoh, Yajima, Torii, Tanaka. The Block Cipher SC2000. Self Evaluation Report. Fujitsu Laboratories LTD and Science University of Tokyo. July 12, 2000.

[3] R.J. Anderson, E. Biham, and L.R. Knudsen. SERPENT - a 128-bit block cipher. A candidate for the Advanced Encryption Standard. Documentation available at `http://www.ii.uib.no/ larsr/serpent`.

[4] 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.

[5] E. Biham. On Matsui's linear cryptanalysis. In A. De Santis, editor, *Advances in Cryptology: EUROCRYPT'94, LNCS 950*, pages 341–355. Springer Verlag, 1995.

[6] E. Biham, A. Biryukov, and A. Shamir. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In J. Stern, editor, *Advances in*

Cryptology: EUROCRYPT'99, LNCS 1592, pages 12–23. Springer Verlag, 1999.

[7] 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.

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

[9] 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.

[10] J. Daemen, L. Knudsen, and V. Rijmen. Linear frameworks for block ciphers. *Design, Codes, and Cryptography.* To appear.

[11] 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.

[12] J. Daemen and V .Rijmen. AES proposal: Rijndael. Submitted as an AES Candidate Algorithm. Available from `http://www.nist.gov/aes`.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] L.R. Knudsen and D. Wagner. Integral cryptanalysis. In preparation, 2001.

[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] 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.

[25] 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.

[26] 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.

[27] 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.

[28] K. Nyberg. Differentially uniform mappings for cryptography. In T. Helleseth, editor, *Advances in Cryptology - EUROCRYPT'93, LNCS 765*, pages 55–64. Springer Verlag, 1993.

[29] 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.

[30] 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.

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

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

[33] Schneier, Kelsey, Whiting, Wagner, Hall, and Ferguson. Twofish: A 128-bit block cipher. Submitted as candidate for AES. Available at `http://www.nist.gov/aes`.

[34] 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.