From: "mrobshaw" <mrobshaw@supanet.com>
To: <AESround2@nist.gov>
Subject: comment
Date: Mon, 15 May 2000 22:23:25 +0100
X-MSMail-Priority: Normal
X-Mailer: Microsoft Outlook Express 5.00.2314.1300
X-MimeOLE: Produced By Microsoft MimeOLE V5.00.2314.1300

Dear Sirs,

Please find attached some comments from Sean Murphy and myself on Twofish.

Yours faithfully,

Matt Robshaw

Differential Cryptanalysis, Key-dependent S-boxes, and Twofish

Sean Murphy[1] and M.J.B. Robshaw[2]

[1] Information Security Group, Royal Holloway, University of London, Egham, Surrey, TW20 0EX. U.K.
sean@dcs.rhbnc.ac.uk
[2] 88 Hadyn Park Road, London, W12 9AG. U.K.
mrobshaw@supanet.com

Abstract. In this note we make some observations on key-dependent S-boxes and differential cryptanalysis. Using basic techniques we give good evidence for the existence of attacks on up to eight rounds of Twofish.

1 Introduction

In this note we summarize some results of a preliminary and basic look at Twofish[4]. In particular we consider the role of key-dependent S-boxes. Key-dependent S-boxes are an interesting device in block cipher design. While intuitively it seems that they must make life very much more difficult for the attacker, it is not clear that this is inevitably the case. In fact, the notion of S-boxes that change from encryption to encryption can be quite useful to an attacker. We sum up the approach in this note as:

> *Instead of choosing the characteristic to fit the S-box, we choose the S-box to fit the characteristic.*

The characteristics in this note were very time-consuming to construct. Time was so short that only a very few options could be tried. Consequently this work should be viewed as illustrative and only a starting point.

2 Some five-round characteristics for Twofish

Here we present two five-round characteristics to Twofish. Throughout this note, the notion of difference is exclusive-or. We only consider the 128-bit key version of Twofish though of course exactly the same type of comments will apply to other key lengths.

2.1 A first five-round characteristic

Here we give the evolution of a five-round characteristic for Twofish. The difference in each 32-bit input word is represented in hexadecimal notation. Full

account is taken of all details in Twofish, including the fixed rotations. The two words to the left in each row are the input differences to the two sets of S-boxes (with the second word being rotated by eight bit positions). The associated probabilities will be discussed in the following sections.

```
80000000  00000000   A0E080A0  AF8FBFAF
                 ↓
00000000  FFFFFFFF   80000000  00000000
                 ↓
00000000  00000000   00000000  FFFFFFFF
                 ↓
00000000  FFFFFFFF   00000000  00000000
                 ↓
40000000  00000000   00000000  FFFFFFFF
                 ↓
50704050  5F1F7F5F   40000000  00000000
```

The evolution of the differential over the third round is clear and holds with probability one. The evolution of the differential in the other rounds is not that obvious and so it is outlined here.

Rounds 2 and 4. The observation that we rely on is the following. Consider the two 32-bit input words to the PHT transformation, say (A, B). The output from the PHT is then ($A+B$, $A+2B$). If we set a difference in (A, B) to be (`0x0`, `0x80000000`) then the difference in the output of the PHT will be (`0x80000000`, `0x0`) with probability one. This difference will also propagate across the additive subkeys that follow, with probability one.

Our aim then is to ensure that the difference in the output from the second MDS matrix is `0x80000000`. It is straightforward to use the inverse of the MDS matrix to give the input difference `0x8CA32FA3`. We notice that all four S-boxes need to be active (a property due to the form of the MDS matrix) yet since the S-boxes are key-dependent we don't know which input differences to consider. However, we can use the property of key-dependent S-boxes to our advantage. In reality it doesn't matter which input difference we use (provided it is non-zero) since there will be some key values that will provide an S-box giving us the characteristic we want, and quite possibly with a reasonable probability.

Since the input form of the difference to the S-boxes doesn't matter, we will choose one that is convenient to us. By choosing `0xFFFFFFFF` we have a difference that is invariant across the single-bit and eight-bit rotations. We further note that the Hamming weight of differences on the input side of the S-boxes is immaterial since they are not involved in any integer addition operations.

In summary, to make characteristics of the form described we have a condition for each S-box; namely that 0xFF $\rightarrow$ 0xA3, 0xFF $\rightarrow$ 0x2F, 0xFF $\rightarrow$ 0xA3, and 0xFF $\rightarrow$ 0x8C for S-boxes 0 through 3 respectively.

Round 1. The input to the round function is of the form (0x80000000, 0x0). This means that there will be one active S-box and due to the MDS matrix there will be four active output bytes. Since we can rely on the key-dependent S-boxes to give us any output difference from the S-box that we like, we can search over all 255 possible non-zero input differences into the MDS matrix to find a useful output. We choose an output that gives a good probability p that the characteristic $(\Delta, 0) \rightarrow (\Delta, \Delta)$ holds over the PHT and integer additions. By choosing Δ = 0xA0E080A0 we accomplish this with p determined experimentally to be 2^{-14} when averaged over random texts and random additive round keys. For Δ = 0xA0E080A0 the input difference to the MDS is 0x80000000 and so we merely require the characteristic 0x80 $\rightarrow$ 0x80 to hold across S-box 3 with non-zero probability in addition to satisfying any other S-box conditions from other rounds.

Round 5. Along similar lines to Round 1, we now need the additional difference 0x40 $\rightarrow$ 0x80 to hold across S-box 3.

All rounds. To derive the five-round differential described we have three conditions on the evolution of a characteristic across S-box 3, and one condition (repeated twice) across S-boxes 0, 1, and 2. A computer search can be used on each S-box individually to identify both the maximum probability for simultaneously satisfying all characteristics across the S-boxes and the number of S-boxes that might provide the necessary differential behavior with a probability above some particular threshold.

For a fraction of (at least) 2^{-40} of the S-boxes the five-round characteristic in Section 2.1 holds with an estimated probability greater than 2^{-53} (maximum 2^{-50}) across the S-boxes and 2^{-28} across the two active sets of PHT and the additive round keys (on average). We would certainly expect there to be some differential effect[3] for this characteristic though the possible extent is unknown.

Note that there is a trade-off between the probability of the characteristic across the S-boxes and the number of keys for which the characteristic will give the stated probability or higher. For this particular case, if we are willing to let the probability of the characteristic across the S-boxes drop from 2^{-53} to 2^{-60}, then experiments suggest that the characteristic will hold for more than 2^{-20} of the keyspace.

[3] A characteristic specifies one particular evolution of differences through the cipher. However a cryptanalyst is typically only concerned with the initial and final difference. Thus the probability of the differential—which is what the crypytanalyst will use—can be significantly higher than the probability of one of the constituent characteristics.

2.2 Another five-round characteristic

Here we show another five-round characteristic for Twofish. It has a different evolution and is of some independent interest.

```
00000000  00000000   00100038  0E000400
                 ↓
0008001C  1C000800   00000000  00000000
                 ↓
00000000  70101060   0008001C  1C000800
                 ↓
0044004E  39001100   00000000  70101060
                 ↓
00000000  00000000   0044004E  39001100
                 ↓
00220027  72002200   00000000  00000000
```

In this case, the evolution of the characteristic over the first and fifth rounds is clear and holds with probability one. Now it is the evolution of the characteristic in the intervening rounds that is not that obvious.

Round 2. Due to the fixed rotations the same inputs will be used to the same S-boxes in round two. This helps reduce the conditions that might be needed on the S-boxes. One choice was to derive the same output from both instances of the MDS matrix; in this case `0x70101060`. There may well be other choices that would be better. (There are at least fifteen easily identified choices that would be useful in similar ways.) Mapping this word back through the inverse of the MDS gives `0x008000C0`. Thus we have our first set of conditions on the S-boxes; `0x1C` $\rightarrow$ `0xC0` for S-box 0 and `0x08` $\rightarrow$ `0x80` for S-box 2. The output from both MDS matrices is `0x70101060` by construction and by experimentation the differential (`0x70101060`, `0x70101060`) $\rightarrow$ (`0x0`, `0x70101060`) holds with probability 2^{-14} on average when computed over random texts and random additive round keys.

Round 3. The input to the round function is of the form (`0x0 0x70101060`). Four S-boxes in this round will be active, and four inactive. Once again we start from the output from the MDS matrices. Clearly one MDS has to give the output `0x0`. We choose the second to have the output `0x00800080`. This choice is particularly interesting because we can use the PHT to our advantage. In order to keep the probability of the characteristic high we aim to keep the number of active S-boxes in the next round low. Yet when we consider the right-hand side of the input difference to this round (`0x0008001C 0xaC000800`) we

see that it will be difficult to get an output from the MDS matrix that doesn't increase the number of active S-boxes in the following round. However across the PHT the characteristic (0x0, 0x00800080) $\rightarrow$ (0x00800080, 0x01000100) holds (experimentally) with average probability 2^{-4} across random texts and random additive round keys. This satisfies our requirements. Mapping 0x00800080 back through the inverse of the MDS matrix gives us 0xC2A3B33F and so we now have another set of conditions on the S-boxes. Namely; 0x70 $\rightarrow$ 0x3F, 0x60 $\rightarrow$ 0xB3, 0x10 $\rightarrow$ 0xA3, 0x10 $\rightarrow$ 0xC2 for S-boxes 0, 1, 2, and 3 respectively.

Round 4. For this round we work backwards. We aim to cancel out the difference on the right of the input to the round. To do this, while taking account of the fixed rotation by one bit position, we need to get as output from the PHT a difference of the form (0x0, 0xE02020C0). This can be accomplished, with probability 2^{-12}, on average, by taking the input to the PHT of the form (0xE02020C0, 0xE02020C0). If we map 0xE02020C0 back through the inverse of the MDS matrix we get 0x006900E9. What is nice about this is that exactly S-boxes 0 and 2 need to be active on the output, and these are exactly the S-boxes active on the input. In short we have the conditions 0x4E $\rightarrow$ 0xE9 and 0x39 $\rightarrow$ 0xE9 for S-box 0 and for S-box 2 we have 0x44 $\rightarrow$ 0x69 and 0x11 $\rightarrow$ 0x69.

All rounds. To get this five-round characteristic we have four conditions on the evolution of a differential across S-boxes 0 and 2, and one condition across S-boxes 1 and 3. In fact the starting values to round 2, 0x002C000C and its rotation were chosen to increase the probability across the S-boxes when these conditions are in place. Many other choices for the bytes 0x2C and 0x0C would have sufficed. Some may lead to much larger classes of keys, even though the maximum probability might be somewhat reduced. There are many different ways of finding such characteristics for Twofish due to the flexibility that key-dependent S-boxes offer the attacker.

For a fraction of (at least) 2^{-42} of the S-boxes this characteristic holds with an estimated probability of between 2^{-63} and 2^{-70} across the S-boxes and 2^{-30} on average across the three active sets of PHT and additive round keys. The extent of any differential effect is unknown.

Again there is a trade-off between the probability of the characteristic across the S-boxes and the number of keys for which the characteristic will give the stated probability or higher. For this particular case, if we are willing to let the probability of the characteristic across the S-boxes drop from 2^{-70} to 2^{-80}, then experiments suggest that the characteristic will hold for more than 2^{-27} of the keyspace.

3 Six-round characteristics for Twofish

The five-round characteristics in Sections 2.1 and 2.2 can be extended to six-round characteristics in the obvious way. In principle these should lead directly to

attacks on seven-round of Twofish. However we take this opportunity to mention a less obvious approach that might also be of some interest.

Consider adding another round to the characteristic of Section 2.1. We see that all eight S-boxes in the following round are going to be active. This would traditionally be viewed as a major problem since either a great many more conditions will be added to those we already have (so the fraction of applicable S-boxes will drop) or the probability of the combined characteristics propagating across the S-boxes will drop.

Instead we observe that four active bytes going into the MDS allows us to have one active byte coming out. Even more important, we can choose that byte to have Hamming weight 1. Thus, we can choose 1024 possible sets of output from the pair of MDS matrices that will have a combined Hamming weight of 2. The hope is that the light differences won't propagate too much across the PHT and the additive subkeys, and that the very light weight of what will therefore become the left side of the text—denoted here by `X` and `Y`—will provide sufficient distinguishing information to launch an attack.

The six-round characteristic would then have the following form where `X` and `Y` will have low Hamming weight.

```
80000000  00000000   A0E080A0  AF8FBFAF
                        ↓
00000000  FFFFFFFF   80000000  00000000
                        ↓
00000000  00000000   00000000  FFFFFFFF
                        ↓
00000000  FFFFFFFF   00000000  00000000
                        ↓
40000000  00000000   00000000  FFFFFFFF
                        ↓
50704050  5F1F7F5F   40000000  00000000
                        ↓
       X          Y   50704050  5F1F7F5F
```

Details

Here we describe in some detail what happens in round six and some of the associated probabilities.

Denote the output from the two MDS matrices as (A, B). All we require is that A and B have Hamming weight one. We can map both A and B back through the MDS matrix using its inverse and this gives us a set of output differences from the two sets of four S-boxes. We know the input differences since they are inherited from the previous round. Therefore we have 1024 sets

of possible differences across the S-boxes in this new round that will be useful to us.

It turns out that for a fraction of (at least) 2^{-20} of the S-boxes, the characteristic as described over six rounds (leading to A and B each of weight one) holds with estimated probability between 2^{-72} and 2^{-85}. In deriving these biases the computation involved characteristics in the first five rounds. By considering differentials improvements can be expected. We also need to account for the probability of crossing two active sets of PHT and the additive subkeys in the first five rounds, and we already know that this takes place with probability 2^{-28} on average.

The final part of the puzzle is to account for what happens to (A, B) as these words pass through the PHT and the additive subkeys. On average, the two words that result (C, D) will have a combined Hamming weight of 9 or less 82% of the time, by experiment. The words (C, D) are exclusive-ored with `0x40000000` and `0x0` and potentially rotated by one bit position.

Clearly there is considerable structure in the output from this six-round characteristic, particularly since the difference on the right-hand side is fully defined over 64-bit bits. It is an open question whether such a technique is useful to a cryptanalyst.

4 Attacks on Reduced-round Twofish

On discovering a characteristic for r rounds of a cipher it is typically prudent to assume that a key-recovery attack on $r + 2$ rounds of the cipher will follow. There has not been time to investigate how the characteristics in this note might be used for that purpose nor to consider the details of key recovery.

The characteristics we have identified in this note can be extended. But to keep within the data requirements and work effort commensurate with the parameters of the AES, account would have to be taken of some of the alternative choices that we haven't had time to examine. Furthermore, the extent of the differential effect would become very important and would be needed if we are to outline a differential attack on eight-round Twofish.

Nevertheless, it seems possible that characteristics and certainly differentials along the lines described in this note could be identified that would compromise eight rounds of Twofish. Of course, this is not in any practical sense, but within the allowable data requirements of the AES. Furthermore, the existence of characteristics that apply to only a fraction of the keyspace leads directly to a class of what might be termed weak keys.

5 Observations

Most observers would agree with Ferguson [1]:

> *The Twofish structure is not easy to analyse. The mixing of various operations makes it hard to give a clean analysis and forces us to use approximation techniques.*

The use of key-dependent S-boxes adds to this complexity and greatly increase the effort required to write automated tools to search for characteristics, differentials and other detailed structure. Despite this, it is possible to make the following observations.

1. It is not clear whether key-dependent S-boxes necessarily offer any additional security over fixed S-boxes. The flexibility of key-dependent S-boxes can actually be used to the advantage of the attacker. The characteristics in this note were constructed by choosing the form of the characteristic across the PHT in a round of Twofish, and then mapping this back through the inverse of the MDS matrix. Provided the same S-boxes were active and inactive on the input and output to the S-box transformation, the actual values of the input and output difference are immaterial. Some S-boxes will give the mapping we need. This gives a lot of flexibility to the cryptanalyst in mounting an attack. The difficulty seems to be mainly one of automating the search to a sufficiently large degree.
2. The fixed rotations by one bit position seem to have a limited impact. Perhaps as we consider very long characteristics they will become a more significant issue for the cryptanalyst. But even this is not clear. At one point the Twofish designers say [4]:

   > *We believe that the one-bit rotations make cryptanalysis harder, if they have any effect at all.*

   But fixed rotations can be can be used by the cryptanalyst to *reduce* the number of active S-boxes in a characteristic. Maybe this is why an alternative view is also supported by one of the Twofish designers [1]:

   > *We have no reason to believe that the 1-bit rotations make Twofish stronger against differential attack.*

3. The fixed rotation by eight bits is intended to lead to conflicts that the cryptanalyst will find hard to resolve. However the use of S-boxes that change with the key mean that there may well be some keys that will resolve any potential conflict.

The characteristics presented in this note were a first attempt. There were many places where different searches could be made with potentially improved results. These basic techniques seem to imply that eight rounds of Twofish offers security commensurate with the parameters of the AES. More sophisticated techniques should yield improved results.

6 Assessing the strength of Twofish

On page 42 of the Twofish report [4] an estimate is made for the number of active S-boxes that might be required in mounting a differential attack on Twofish. This was done by considering the number of active S-boxes that might be needed for a 12-round characteristic to Twofish.

Using this approach of counting the number of active S-boxes, it was claimed that a 12-round characteristic could be constructed with 20 active S-boxes. It was further stated that to be prudent one should perhaps expect a 3R-attack, i.e. one on 12+3 rounds of Twofish. It exemplifies the typical approach to assessing the security of a cipher. This reasoning was followed up in a later technical note and additional analysis suggested that a conservative estimate for a differential attack on 15 rounds of Twofish would require around 2^{103} chosen plaintexts. Of course this estimate is likely to be an over-simplification. But it is a very appropriate way to provide conservative bounds to the security of a cipher.

7 Conclusions

The basic observations in this note do not pose a threat to Twofish with 16 rounds. However they question the security of eight-round Twofish.

The unusual design of Twofish means that unusual techniques are likely to be required in any successful attempt at cryptanalysis. This certainly makes starting out a whole lot harder. Beyond that, Twofish remains untested.

References

1. N. Ferguson. Upper bounds on differential characteristics in Twofish. Counterpane Systems. August 17, 1998.
2. N. Ferguson. Impossible differentials in Twofish. Counterpane Systems. October 19, 1999.
3. J. Kelsey. Key separation in Twofish. Counterpane Systems. April 7, 2000.
4. B. Schneier, J. Kelsey, D. Whiting, D. Wagner, C. Hall, and N. Ferguson. Twofish: A 128-bit Block Cipher. 15 June, 1998.

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