Security of Camellia against Truncated Differential Cryptanalysis

Masayuki Kanda[1,2] and Tsutomu Matsumoto[2]

[1] NTT Information Sharing Platform Laboratories
1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan
kanda@isl.ntt.co.jp
[2] Yokohama National University
79-5 Tokiwadai, Hodogaya-ku, Yokohama-shi, Kanagawa, 240-8501 Japan
tsutomu@mlab.jks.ynu.ac.jp

Abstract. This paper studies security against truncated differential cryptanalysis from the "designer's" standpoint. In estimating the security, we use the upper bound of truncated differential probability. Previous works, Knudsen, Matsui and Moriai et al., searched for effective truncated differentials to attack byte-oriented block ciphers and computed the exact probability of the differentials. In this paper, we discuss the following items from the designer's standpoint; (a) truncated differential probability of effective active-s-box, (b) XOR cancellation probability, and (c) effect of auxiliary functions, e.g., FL/FL^{-1}-functions. We then combine them with Matsui's search algorithm and evaluate the security of Camellia, jointly developed by NTT and Mitsubishi Electric Corporation, against truncated differential cryptanalysis. We prove (from the designer's standpoint) that variants of Camellia with more than 11 rounds are secure against truncated differential cryptanalysis even if weak-key FL/FL^{-1}-functions are taken into consideration.

1 Introduction

At the second FSE in 1994, Knudsen proposed truncated differential cryptanalysis [K95], as an extension of differential cryptanalysis [BS93]. He defined truncated differentials as differentials where only a part of the difference in the ciphertexts (after a number of rounds) can be predicated. His first concept of truncated differentials was quite wide. Generally speaking, truncated differentials are now regarded as subsets of the characteristics that can be used to attack the cipher by using information on whether several bits of the difference are zero or not. In particular, "bytewise" truncated differentials, where one byte of the difference is regarded as 1 (non-zero) or 0 (zero), are useful in attacking byte-oriented block ciphers. For example, Knudsen and Berson attacked 5-round SAFER [KB96], and Matsui, Tokita [MT99,M99] and Moriai et al. [MSA+99] attacked the reduced version of E2 [KMA+00]. This means that we[1] have designed Camellia [AIK+00], jointly developed by NTT and Mitsubishi Electric

[1] Kanda is a member of the Camellia design team.

M. Matsui (Ed.): FSE 2001, LNCS 2355, pp. 286–299, 2002.

Corporation, while considering its immunity to bytewise truncated differential cryptanalysis, since Camellia is a byte-oriented block cipher.

It is well known that there are two ways to estimate the security of ciphers against cryptanalyses; i.e., security measures from the "attacker's standpoint" and those from the "designer's standpoint." For example, in the case of differential cryptanalysis, one former measure is the maximum differential characteristic probability, and one of the latter is the upper bound of differential characteristic probability. Researchers who intend to attack ciphers have to find effective characteristic(s) and evaluate the exact characteristic probability from the attacker's standpoint in order to estimate the success rate of the attack and the computational load. On the other hand, it is sufficient for designers to show that the upper bound of the probability is low enough from the designer's standpoint in proving that the cipher is secure against differential cryptanalysis. This is because in such a case, all characteristics are useless in attacking the cipher, and it is unnecessary to calculate the stricter characteristic probability. Indeed, many recent ciphers have been designed on the basis of the upper bound of characteristic probability, such as SHARK [RDP$^+$96], Square [DKR97], Rijndael [DR98], SERPENT [ABK98], E2, Camellia, and MISTY [M97].

In this paper, we study security against truncated differential cryptanalysis from the designer's standpoint. Our evaluation is based on the upper bound of truncated differential probability.

First, we discuss the following items.

- How do effective active-s-boxes reduce truncated differential probability? Knudsen, Matsui, and Moriai et al. regarded 8-bit effective active-s-boxes as reducing the probability by 2^{-8}, but is this right (from the designer's viewpoint)?
- How do XOR cancellations reduce truncated differential probability? Matsui and Moriai et al. regarded them as reducing the probability by $(2^{-8})^h$, where h denotes the number of bytewise XOR cancellations. Is this right (from the designer's viewpoint)?
- How do auxiliary functions, e.g., FL/FL^{-1}-functions, revise truncated differential probability?

We then combine the above results with Matsui's search algorithm [M99] and evaluate the security of Camellia against truncated differential cryptanalysis.

This paper is organized as follows. In Sect. 2, we introduce some notations and definitions. The above items are discussed in Sect. 3 to Sect. 5, respectively. Finally, we show the security of Camellia against truncated differential cryptanalysis in Sect. 6, and conclude with a summary in Sect. 7.

2 Preliminaries

2.1 Notations

- $X = (x_1, \ldots, x_m),\ x_i \in \mathbb{Z}_2^n,\ (i = 1, \ldots, m)$: vector X over GF$(2^n)^m$.
- $\Delta X = (\Delta x_1, \ldots, \Delta x_m),\ \Delta x_i \in \mathbb{Z}_2^n,\ (i = 1, \ldots, m)$: difference of X.

- $\delta X = (\delta x_1, \ldots, \delta x_m),\ \delta x_i \in \mathbb{Z}_2,\ (i = 1, \ldots, m)$: truncated differential of X.
- $\#\{S\}$: A number of elements in set S.

2.2 Definitions

Definition 1. *An active s-box is defined as an s-box given a non-zero input difference. When the s-box is bijective, it yields a non-zero output difference.*

Definition 2. *An Effective active-s-box is defined as an active s-box which outputs some specified difference; i.e., the output difference is related to the output difference(s) of other s-box(es). On the other hand, an active s-box which can output any non-zero difference is called a non-effective active-s-box.*

Definition 3. *When two non-zero differences are input to an XOR (exclusive-OR) operation but it outputs a zero difference, the XOR operation is called an XOR cancellation.*

Definition 4. *For any given Δx, $\Delta z \in \mathbb{Z}_2^n$, the differential probability of s-box $p_s(\Delta x \to \Delta z)$ is defined as:*

$$\begin{aligned} p_s(\Delta x \to \Delta z) &= \Pr_{x \in \mathbb{Z}_2^n}[s(x) \oplus s(x \oplus \Delta x) = \Delta z] \\ &= \frac{\#\{x \in \mathbb{Z}_2^n | s(x) \oplus s(x \oplus \Delta x) = \Delta z\}}{2^n} \end{aligned}$$

Definition 5. *For any given ΔX, $\Delta Y \in (\mathbb{Z}_2^n)^m$, the differential probability of F-function $p_F(\Delta X \to \Delta Y)$ is defined as:*

$$\begin{aligned} p_F(\Delta X \to \Delta Y) &= \Pr_{x \in (\mathbb{Z}_2^n)^m}[F(X) \oplus F(X \oplus \Delta X) = \Delta Y] \\ &= \frac{\#\{X \in (\mathbb{Z}_2^n)^m | F(X) \oplus F(X \oplus \Delta X) = \Delta Y\}}{(2^n)^m} \end{aligned}$$

Definition 6. *For any given $\Delta x \in \mathbb{Z}_2^n$, a function $\chi : \mathrm{GF}(2^n) \to \mathrm{GF}(2)$ is defined as:*

$$\chi(\Delta x) = \delta x = \begin{cases} 0 & \textit{if } \Delta x = 0 \\ 1 & \textit{if } \Delta x \neq 0 \end{cases}$$

Furthermore, for any given $\Delta X \in (\mathbb{Z}_2^n)^m$, $\chi(\Delta X) = \delta X = (\chi(\Delta x_1), \ldots, \chi(\Delta x_m))$.

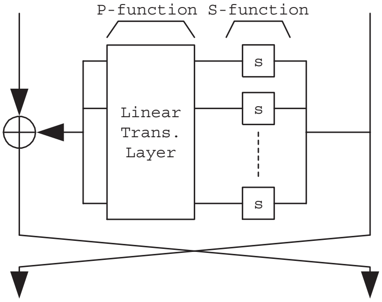

Fig. 1. Model

Definition 7. *For any given $\Delta X, \Delta Y \in \mathrm{GF}(2^n)^m$ and $\delta X, \delta Y \in \mathrm{GF}(2)^m$, the truncated differential probability of F-function $p_F(\delta X \to \delta Y)$ is defined as:*

$$p_F(\delta X \to \delta Y) = \frac{\displaystyle\sum_{\substack{\chi(\Delta X) = \delta X \\ \chi(\Delta Y) = \delta Y}} \Pr_{X \in (\mathbb{Z}_2^n)^m}[F(X) \oplus F(X \oplus \Delta X) = \Delta Y]}{\#\{\Delta X | \chi(\Delta X) = \delta X\}}$$

The following useful theorem can be obtained easily from Definition 7.

Theorem 1. *The upper bound of the truncated differential probability of F-function $p_F(\delta X \to \delta Y)$ is denoted as:*

$$\max_{\substack{\Delta X \\ \chi(\Delta X) = \delta X}} \sum_{\chi(\Delta Y)=\delta Y} \Pr_{X \in (\mathbb{Z}_2^n)^m}[F(X) \oplus F(X \oplus \Delta X) = \Delta Y]$$

Theorem 1 means that the upper bound of truncated differential probability $p_F(\delta X \to \delta Y)$ is obtained from the case that some input difference ΔX, which satisfies $\chi(\Delta X) = \delta X$, yields the maximum value of the above equation. Conversely, in order to prove immunity to truncated differential cryptanalysis, the designer should consider the upper bound of the probability assuming that difference ΔX which satisfies $\chi(\Delta X) = \delta X$ and yields the maximum value of the above equation is input to the F-function.

2.3 Model

Throughout this paper, we consider Feistel ciphers with SPN round function (See Fig. 1). A nonlinear layer that consists of m parallel s-boxes is called S-function and a linear transformation layer is called P-function. Hereafter, we assume that the s-boxes are 8-bit bijective substitution tables, since this eases the discussion.

Note that, since we assume that a round key, which is input to one round function, is generated independently and uniform randomly, and is bitwise XORed with input data, we ignore the effect of the round key throughout this paper.

3 Impact of Effective Active-s-Boxes

In this section, we study the effect of effective active-s-boxes.

As mentioned in Definition 1, an active s-box transforms a non-zero input difference to a non-zero output difference with some differential probability computed by Definition 4 because the s-box is bijective. Here, since the s-box is a nonlinear transformation function, the differential probability is *not* one in general. That is, $p_s(\Delta x \to \Delta z) < 1$ for $\Delta x \neq 0$ and $\Delta z \neq 0$. This means that an active s-box always reduces the probability of differential characteristics. Accordingly, one security measure against differential cryptanalysis is based on the minimum number of active s-boxes in the cipher, which shows the upper bound of differential characteristic probability.

A truncated differential just shows whether some bytewise difference is nonzero or zero, while the bytewise difference itself has some non-zero value. Thus, an active s-box that is bijective *always* transforms a non-zero bytewise input truncated differential to a non-zero bytewise output truncated differential; i.e., $p_s(\delta x \to \delta z) = 1$ for $\delta x \neq 0$ and $\delta z \neq 0$. Such an active s-box is a non-effective active-s-box, see Definition 2, since it does *not* reduce the probability of truncated differentials. Accordingly, when each active s-box outputs a nonzero output difference independently, all active s-boxes are non-effective active-s-boxes and do not reduce any truncated differential probability.

On the other hand, how do active s-boxes reduce the truncated differential probability if they output some dependent output difference(s)? Here, a dependent output difference means the output difference that is determined by output difference(s) of other active s-box(es). In this case, it is necessary to consider the output difference(s) itself of the active s-box(es) and not the output truncated differential(s). Such an active s-box is an effective active-s-box mentioned in Definition 2, and it reduces the probability of truncated differentials in the same way as an active s-box in differential cryptanalysis.

Hereafter, we show how effective active-s-boxes reduce the truncated differential probability.

It is well known that the inverse function in GF(2^8) provides the minimum value of the maximum differential probability, i.e., 2^{-6}. Thus, many ciphers use s-boxes based on the inverse function in GF(2^8). Here, assume that all s-boxes consist of combinations of the inverse function in GF(2^8) and an affine transformation. $\delta X, \delta Z$ denote an input truncated differential and an output truncated differential of S-function, respectively.

Theorem 2. *Assume that all s-boxes consist of combinations of the inverse function in* GF(2^8) *and an affine transformation. Let the number of active s-boxes be k ($k \geq 2$). When all active s-boxes output the same output difference*[2]*, the truncated differential probability of S-function $p_S(\delta X \to \delta Z)$ satisfies the following relationship.*

$$p_S(\delta X \to \delta Z) \leq 2^{-7(k-1)}(1 + 2^{k-7} - 2^{-6})$$

[2] In this case, the number of effective active-s-boxes is $k - 1$.

Proof. It is easily confirmed that the difference distribution table of the inverse function in GF(2^8) has one 4-possible pairs and 126 2-possible pairs in each column and each row. Furthermore, an affine transformation only interchanges the columns of the difference distribution table. Thus, for each Δz, the differential probability of s-box $p_s(\Delta x \rightarrow \Delta z)$ is 2^{-6} for one entry and 2^{-7} for 126 entries from 256 entries of Δx, since the s-box consists of combinations of the inverse function in GF(2^8) and an affine transformation.

Accordingly, in order to calculate the upper bounds of $p_S(\delta X \rightarrow \delta Z)$ when all active s-boxes output the same output difference Δz, we should consider that the same input difference Δx is input to all s-boxes. This leads to the following relationship.

$$p_S(\delta X \rightarrow \delta Z) \leq \max_{\Delta x \neq 0} \sum_{\Delta z \neq 0} \{p_s(\Delta x \rightarrow \Delta z)\}^k = (2^{-6})^k + 126 \cdot (2^{-7})^k$$

Q.E.D.

Theorem 2 can easily be extend for the generic case as follows.

Theorem 3. *Assume that each row (Δx) of the difference distribution table of an 8-bit s-box has n_i 2^{8-i}-possible pairs ($i = 0, 1, \ldots, 7$)[3]. Let the number of active s-boxes be k ($k \geq 2$). When all active s-boxes output the same output difference[4], the truncated differential probability of S-function $p_S(\delta X \rightarrow \delta Z)$ satisfies the following relationship.*

$$p_S(\delta X \rightarrow \delta Z) \leq \max_{\Delta x \neq 0} \left\{ 2^{-7(k-1)} + \sum_{i=0}^{6} \frac{n_i}{2^i} \left(2^{-i(k-1)} - 2^{-7(k-1)} \right) \right\}$$

Proof. This theorem can be obtained easily from the following two equations.

$$\sum_{\Delta z \neq 0} \{p_s(\Delta x \rightarrow \Delta z)\}^k = \sum_{i=0}^{7} n_i \cdot 2^{-ik}, \quad \sum_{i=0}^{7} n_i \cdot 2^{8-i} = 256$$

Q.E.D.

Theorem 4. *Assume that all s-boxes consist of combinations of the inverse function in* GF(2^8) *and an affine transformation. Let the number of active s-boxes be 3. Assume that an active s-box outputs the output difference denoted by the XOR of two other output differences[5]. That is, these output differences represent $(\alpha, \beta, \alpha \oplus \beta), \alpha \neq \beta$. Then, the truncated differential probability of S-function $p_S(\delta X \rightarrow \delta Z)$ satisfies the following relationship.*

$$p_S(\delta X \rightarrow \delta Z) \leq 2^{-7} + 2^{-14} + 2^{-20} \simeq 2^{-7}$$

[3] In the case of Theorem 2, $n_0 = n_1 = n_2 = n_3 = n_4 = n_5 = 0, n_6 = 1, n_7 = 126$.

[4] In this case, the number of effective active-s-boxes is $k - 1$.

[5] In this case, the number of effective active-s-boxes is 1.

Proof. As mentioned at the proof of Theorem 2, for each Δz, the differential probability of s-box $p_s(\Delta x \rightarrow \Delta z)$ is 2^{-6} for one entry and 2^{-7} for 126 entries from 256 entries of Δx. Also, for each Δx, the differential probability is 2^{-6} for one entry and 2^{-7} for 126 entries from 256 entries of Δz.

Let $\Delta x_1, \Delta x_2, \Delta x_3$ denote input differences to active s-boxes. Without loss of generality, we assume that $p_s(\Delta x_1 \rightarrow \alpha_0) = 2^{-6}$, $p_s(\Delta x_2 \rightarrow \beta_0) = 2^{-6}$, $p_s(\Delta x_3 \rightarrow \alpha_0 \oplus \beta_0) = 2^{-6}$, and $\alpha_0 \neq \beta_0$. The differential probability and the entries of such probability are determined as follows by combinations of the output differences $(\alpha, \beta, \alpha \oplus \beta)$.

1. $(\Delta x_1, \Delta x_2, \Delta x_3) \mapsto (\alpha_0, \beta_0, \alpha_0 \oplus \beta_0)$: the differential probability is $(2^{-6})^3$ and the entry is 1.
2. $(\Delta x_1, \Delta x_2, \Delta x_3) \mapsto (\alpha_0, \beta, \alpha_0 \oplus \beta)$, $\beta \neq \beta_0$: the differential probability is $2^{-6} \cdot (2^{-7})^2$ and the entries are (at most) 126.
3. $(\Delta x_1, \Delta x_2, \Delta x_3) \mapsto (\alpha, \beta_0, \alpha \oplus \beta_0)$, $\alpha \neq \alpha_0$: the differential probability is $2^{-6} \cdot (2^{-7})^2$ and the entries are (at most) 126.
4. $(\Delta x_1, \Delta x_2, \Delta x_3) \mapsto (\alpha, \alpha \oplus \alpha_0 \oplus \beta_0, \alpha_0 \oplus \beta_0)$, $\alpha \neq \alpha_0$: the differential probability is $2^{-6} \cdot (2^{-7})^2$ and the entries are (at most) 126.
5. $(\Delta x_1, \Delta x_2, \Delta x_3) \mapsto (\alpha, \beta, \alpha \oplus \beta)$, $\alpha \neq \alpha_0$, $\beta \neq \beta_0$, $\alpha \oplus \beta \neq \alpha_0 \oplus \beta_0$: the differential probability is $(2^{-7})^3$ and the entries are (at most) $126 \cdot 125$.

From the above, Theorem 1 leads to the following relationship.

$$\begin{aligned} p_S(\delta X \rightarrow \delta Z) &\leq \max_{\substack{(\Delta x_1, \Delta x_2, \Delta x_3) \\ \chi(\Delta X) = \delta X}} \sum_{\substack{(\alpha_0, \beta_0), (\alpha_0, \beta), \\ (\alpha, \beta_0), (\alpha, \beta)}} p_S(\Delta X \rightarrow \Delta Z) \\ &= 2^{-18} + 126 \cdot 2^{-20} + 126 \cdot 2^{-20} + 126 \cdot 2^{-20} + 126 \cdot 125 \cdot 2^{-21} \\ &= 2^{-7} + 2^{-14} + 2^{-20} \end{aligned}$$

Q.E.D.

Theorem 5. *Assume that all s-boxes consist of combinations of the inverse function in* GF(2^8) *and an affine transformation. Let the number of active s-boxes be 7. Assume that seven output differences represent* $(\alpha, \alpha, \alpha, \beta, \beta, \alpha \oplus \beta, \alpha \oplus \beta)$[6]. *Then, the truncated differential probability of S-function* $p_S(\delta X \rightarrow \delta Z)$ *satisfies the following relationship.*

$$p_S(\delta X \rightarrow \delta Z) \leq 2^{-35} + 2^{-38} - 2^{-40} - 2^{-44} + 3 \cdot 2^{-48} \simeq (2^{-7})^5$$

Since the proof of this theorem is very similar to the proof of Theorem 4, the proof is omitted here. From Theorem 2, Theorem 4, and Theorem 5, we can obtain the following conjecture.

Conjecture 1. *Assume that all s-boxes consist of combinations of the inverse function in* GF(2^8) *and an affine transformation. The upper bound of truncated differential probability is reduced by approximately* 2^{-7} *per effective active-s-box*[7].

[6] In this case, the number of effective active-s-boxes is 5.

[7] It is expected that the stricter upper bound is $2^{-6.9}$.

Knudsen, Matsui and Moriai et al. estimated that the truncated differential probability is reduced by approximately 2^{-8} (exactly $\frac{1}{255}$) per effective active-s-box assuming that the output difference distribution of the s-box is uniform. However, since the distribution of an even-bit s-box is not uniform, designers should not ignore the case when truncated differential probability is reduced by approximately 2^{-7} per effective active-s-box. Accordingly, we regard an effective active-s-box as reducing the probability by 2^{-7} to evaluate the security of Camellia against truncated differential cryptanalysis.

4 XOR Cancellation Probability

Let $\Delta X^{(r-1)}$, $\delta X^{(r-1)}$ be an input difference and an input truncated differential in the $(r-1)$-th round, respectively. Let $\Delta Y^{(r)}$, $\delta Y^{(r)}$ be an output difference and an output truncated differential in the r-th round, respectively. Here, $(\Delta X^{(0)}, \Delta X^{(1)})$ denotes a difference of plaintexts.

$$\Delta X^{(r-1)} = (\Delta x_1^{(r-1)}, \ldots, \Delta x_m^{(r-1)}) \ , \ \delta X^{(r-1)} = (\delta x_1^{(r-1)}, \ldots, \delta x_m^{(r-1)})$$
$$\Delta Y^{(r)} = (\Delta y_1^{(r)}, \ldots, \Delta y_m^{(r)}) \ , \ \delta Y^{(r)} = (\delta y_1^{(r)}, \ldots, \delta y_m^{(r)})$$

An input difference $\Delta X^{(r+1)}$ and an input truncated differential $\delta X^{(r+1)}$ in the $(r+1)$-th round are obtained as follows.

$$\Delta X^{(r+1)} = (\Delta x_1^{(r-1)} \oplus \Delta y_1^{(r)}, \ldots, \Delta x_m^{(r-1)} \oplus \Delta y_m^{(r)})$$
$$\delta X^{(r+1)} = (\delta x_1^{(r-1)} \oplus \delta y_1^{(r)}, \ldots, \delta x_m^{(r-1)} \oplus \delta y_m^{(r)})$$

Since an XOR cancellation occurs when $\Delta x_i^{(r-1)} = \Delta y_i^{(r)} (\neq 0)$, XOR cancellation probability (of byte operation) is approximately 2^{-8} (exactly $\frac{1}{255}$) if $\Delta x_i^{(r-1)}$ and/or $\Delta y_i^{(r)}$ is determined uniform randomly. Thus, in the previous estimation, the XOR cancellation probability of truncated differentials denotes $(2^{-8})^h$, where h is the number of XOR cancellations. Here, $h = hw(\delta X^{(r-1)} \vee \delta Y^{(r)}) - hw(\delta X^{(r+1)})$ $(\leq hw(\delta X^{(r-1)} \wedge \delta Y^{(r)}))$, and $hw(X)$ denotes the Hamming weight of X.

Note that, when the XOR cancellation probability is satisfied as mentioned above, there is a tacit assumption that no relation exists between $\Delta x_i^{(r-1)}$ and $\Delta x_j^{(r-1)}$ $(i \neq j)$ or between $\Delta y_i^{(r)}$ and $\Delta y_j^{(r)}$ $(i \neq j)$. Generally speaking, it is expected that there is no relation between $\Delta x_i^{(r-1)}$ and $\Delta x_j^{(r-1)}$ $(i \neq j)$ if $(r-1)$ is large. However, this cannot be expected if $(r-1)$ is small. Accordingly, XOR cancellation probability is dependent on whether there is a relation between $\Delta y_i^{(r)}$ and $\Delta y_j^{(r)}$ $(i \neq j)$ or not.

For example, consider E2 and Camellia as shown in Fig. 2. Since an S-P-S structure is applied to the round function of E2, output difference Δy_i is represented as follows.

$$\Delta y_i = s(f_i(s(\Delta x_1), \ldots, s(\Delta x_8))), \ (i = 1, \ldots, 8),$$

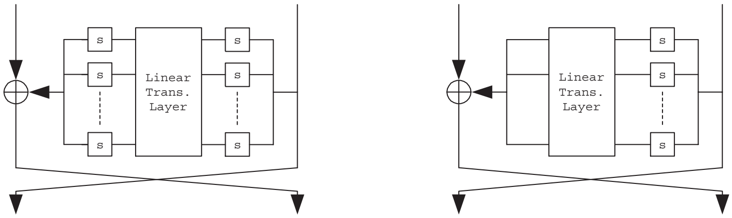

Fig. 2. Round Functions of E2 (left hand) and Camellia (right hand)

where $s()$ denotes an s-box and $f_i()$ shows a linear transformation. It turns out that, for any input difference ΔX belonging to the given input truncated differential δX, there is no relation between Δy_i and Δy_j $(i \neq j)$ even if $f_i(s(\Delta x_1), \ldots, s(\Delta x_8)) = f_j(s(\Delta x_1), \ldots, s(\Delta x_8))$ for some i, j. That is, all output differences Δy_i are determined independently. Accordingly, as in the previous estimation, the XOR cancellation probability of E2 can be represented as $(2^{-8})^h$, where h denotes the number of bytewise XOR cancellations.

On the other hand, since an S-P structure is applied to the round function of Camellia, output difference Δy_i is represented as follows.

$$\Delta y_i = f_i(s(\Delta x_1), \ldots, s(\Delta x_8)),\ (i = 1, \ldots, 8).$$

More concretely, we can rewrite as follows.

$$\begin{aligned}
\Delta y_1 &= s(\Delta x_1) \oplus s(\Delta x_3) \oplus s(\Delta x_4) \oplus s(\Delta x_6) \oplus s(\Delta x_7) \oplus s(\Delta x_8)\\
\Delta y_2 &= s(\Delta x_1) \oplus s(\Delta x_2) \oplus s(\Delta x_4) \oplus s(\Delta x_5) \oplus s(\Delta x_7) \oplus s(\Delta x_8)\\
\Delta y_3 &= s(\Delta x_1) \oplus s(\Delta x_2) \oplus s(\Delta x_3) \oplus s(\Delta x_5) \oplus s(\Delta x_6) \oplus s(\Delta x_8)\\
\Delta y_4 &= s(\Delta x_2) \oplus s(\Delta x_3) \oplus s(\Delta x_4) \oplus s(\Delta x_5) \oplus s(\Delta x_6) \oplus s(\Delta x_7)\\
\Delta y_5 &= s(\Delta x_1) \oplus s(\Delta x_2) \oplus s(\Delta x_6) \oplus s(\Delta x_7) \oplus s(\Delta x_8)\\
\Delta y_6 &= s(\Delta x_2) \oplus s(\Delta x_3) \oplus s(\Delta x_5) \oplus s(\Delta x_7) \oplus s(\Delta x_8)\\
\Delta y_7 &= s(\Delta x_3) \oplus s(\Delta x_4) \oplus s(\Delta x_5) \oplus s(\Delta x_6) \oplus s(\Delta x_8)\\
\Delta y_8 &= s(\Delta x_1) \oplus s(\Delta x_4) \oplus s(\Delta x_5) \oplus s(\Delta x_6) \oplus s(\Delta x_7)
\end{aligned}$$

Now, let an input truncated differential be $\delta X = (10001000)$. The above equations are represented as;

$$\begin{aligned}
&\Delta y_1 = \Delta y_5 = s(\Delta x_1),\\
&\Delta y_2 = \Delta y_3 = \Delta y_8 = s(\Delta x_1) \oplus s(\Delta x_5),\\
&\Delta y_4 = \Delta y_6 = \Delta y_7 = s(\Delta x_5).
\end{aligned}$$

This means that the relation between Δy_1 and Δy_5 is always satisfied regardless of input difference Δx_1. so also are the relations between Δy_2, Δy_3 and Δy_8, and between Δy_4, Δy_6 and Δy_7.

Let $\delta X^{(r)} = (10001000)$, $\delta X^{(r-1)} = (01100001)$, and $\Delta x_2^{(r-1)} = \Delta x_3^{(r-1)} = \Delta x_8^{(r-1)} = \alpha$, for example. If an input truncated differential of the $(r+1)$-th round satisfies $\delta X^{(r+1)} = (10011110)$, then the XOR cancellation probability is 2^{-8}, not $(2^{-8})^3$, since it is sufficient to satisfy $s(\Delta x_1) \oplus s(\Delta x_5) = \alpha$. Accordingly, we obtain the following theorem.

Theorem 6. *Let input truncated differentials of the $(r-1)$-th round and the $(r+1)$-th round be $\delta X^{(r-1)}$ and $\delta X^{(r+1)}$, respectively. Also, let an output truncated differential and an output difference of the r-th round be $\delta Y^{(r)}$ and $\Delta Y^{(r)}$, respectively. Furthermore, an (i,j) element of $m \times m$ matrix $\mathcal{D}$ is determined by;*

$$d_{ii} = \begin{cases} 1 & \text{if} \quad \delta x_i^{(r-1)} \cdot \delta y_i^{(r)} = 1 \cap \delta x_i^{(r+1)} = 0 \\ 0 & \text{otherwise} \end{cases}$$
$$d_{ij} = 0 \qquad \text{if} \quad i \neq j$$

$\mathcal{D} \cdot \Delta Y^{(r)}$ picks up the value(s) which leads to XOR cancellation. The same value of $\mathcal{D} \cdot \Delta Y^{(r)}$ means that there is a relation between the output differences. Accordingly, XOR cancellation probability is represented as approximately $(2^{-8})^k$, where k is the rank of $\mathcal{D} \cdot \Delta Y^{(r)}$. Note that, since S-P-S round function always makes the rank $k = h$, the previous estimation is a special case of this theorem.

5 Impact of Auxiliary Functions

Auxiliary functions, which are heterogeneous functions from round function (core structure), are inserted before and after the core and/or at the inside of the core. Generally speaking, it is expected that auxiliary functions improve the security. Thus, if the core has sufficient immunity to cryptanalyses, it is considered that the cipher, which has the core and auxiliary functions, is also secure. However, if an auxiliary function is a key-dependent linear transformation such as the FL/FL^{-1}-function of Camellia, there is some possibility of generating weak-key auxiliary functions. That is, whether an auxiliary function decreases the security depends on the key used. This means that designers should estimate the security of the cipher assuming the use of weak-key auxiliary functions.

In this section, we study an effect on security when inserting key-dependent linear transformation functions as auxiliary functions. Here, we focus on the FL-function of Camellia. The differential transition of the FL-function is described as follows.

$$\begin{aligned} FL : {} & \mathrm{GF}(2^{32})^4 \to \mathrm{GF}(2^{32})^2; \\ & (\Delta A, \Delta B, k_1, k_2) \mapsto (\Delta C, \Delta D), \\ & \Delta D = ((\Delta A \wedge k_1) <<< 1) \oplus \Delta B), \\ & \Delta C = \Delta A \oplus (\Delta D \wedge \overline{k_2}), \end{aligned}$$

where $<<<$ 1 denotes one-bit right circular rotation, k_1, k_2 are subkeys, and $\overline{k_2}$ represents the complement of k_2. Thus, through the FL-function, an input difference of the FL-function $(\Delta A, \Delta B)$ is mapped to some output difference $(\Delta C, \Delta D)$ by;

$$(\Delta A, \Delta B) = (0,0) \overset{k_1,k_2}{\longmapsto} (\Delta C, \Delta D) = (0,0),$$
$$(\Delta A, \Delta B) \neq (0,0) \overset{k_1,k_2}{\longmapsto} (\Delta C, \Delta D) \in \mathrm{GF}(2^{32})^2 \backslash (0,0).$$

Since $(\Delta A, \Delta B) \neq (0,0)$ is transformed to some $(\Delta C, \Delta D) \in \mathrm{GF}(2^{32})^2 \backslash (0,0)$, there is some possibility of transforming $(\Delta A', \Delta B')$ with relatively high differential probability to $(\Delta C, \Delta D) \neq (0,0)$, even if $(\Delta C, \Delta D) \neq (0,0)$ has low differential probability before applying a weak-key FL-function. This means that taking the weak-key FL-function into consideration is equivalent to regarding the probability of all non-zero output differences of the FL-function as the maximum probability of non-zero input difference of the FL-function.

The above statement is applicable to the FL^{-1}-function of Camellia. Accordingly, we define the effect of FL/FL^{-1}-function as follows.

Definition 8. *Let δI, δO be an input truncated differential and an output truncated differential of FL/FL^{-1}-function, respectively. The effect of FL/FL^{-1}-function is that the probability of non-zero truncated differentials is changed to;*

$$\forall \delta O \neq 0, \quad \Pr((\delta X(0), \delta X(1)) \to \delta O) = \max_{\delta I \neq 0} \Pr((\delta X(0), \delta X(1)) \to \delta I),$$

where $(\delta X(0), \delta X(1))$ is a truncated differential of plaintexts.

6 Upper Bound of Best Bytewise Characteristics of Camellia

In this section, we look for the upper bound of best bytewise characteristics of Camellia. Our search algorithm is based on Matsui's search algorithm (width first basis) [M99], and we modify his algorithm at the following points.

Step 1: Generation of Truncated Differential Probability of *F*-Function

In [M99], since he assumed that a difference distribution table of s-box was uniform, he regarded the truncated differential probability as reducing 2^{-8} per effective active-s-box. As mentioned in Sect. 3, however, the difference distribution table of the s-box which consists of combinations of the inverse function in $\mathrm{GF}(2^8)$ and an affine transformation is not uniform, and the upper bound of truncated differential probability is reduced by approximately 2^{-7} per effective active-s-box. Thus, we regard the upper bound of truncated differential probability of F-function as being 2^{-7r}, where r is the number of effective active-s-boxes of F-function. r is obtained as follows for given input truncated differential δX and output truncated differential δY.

$$r = rank(\mathcal{F}(\overline{\delta Y}) P \delta X),$$

where $\overline{\delta Y}$ is the complement of δY, P denotes the matrix of the linear transformation in GF(2), and $\mathcal{F}(\overline{\delta Y})$ represents the 8×8 diagonal matrix whose (i,i) component equals $\overline{\delta y_i}$ for $i = 1, \ldots, 8$.

Step 2: Generation of Truncated Differential Probability of Round Function

In [M99], since he assumed that there was no relation among bytewise output differences of F-function, he regarded the XOR cancellation probability as being 2^{-8n}, where n is the number of XOR cancellation. As mentioned in Sect. 4, however, there are some relations among bytewise output differences of F-function, since the F-function of Camellia consists of S-P structure. Thus, Theorem 6 makes XOR cancellation probability 2^{-8k}, where k is the rank of $\mathcal{D} \cdot \Delta Y$.

The simple calculation on the rank of $\mathcal{D} \cdot \Delta Y$ is shown below.

Definition 9. *Let A, B, C and D denote 4-bit independent difference patterns, i.e, $A = 0001, B = 0010, C = 0100,$ and $D = 1000$. Furthermore, assume that a given output difference $\Delta Z = (\Delta z_1, \ldots, \Delta z_8)$ can be labeled by combinations of the difference patterns $A, B, C,$ and D. Rank representation of ΔZ is defined as the description of how each Deltaz$_i$$(\neq 0)$, $(i = 1, \ldots, 8)$ is represented by a combination of $A, B, C,$ and D.*

For example, let a difference ΔZ be $\Delta Z = (\Delta z_1, \Delta z_2, 0, 0, \Delta z_5, 0, \Delta z_7, \Delta z_8)$, and the three following relations exist; $\Delta z_1 = \Delta z_5$, $\Delta z_2 = \Delta z_8$, $\Delta z_7 = \Delta z_1 \oplus \Delta z_2$. When difference patterns are labeled as $A = \Delta z_1$ and $B = \Delta z_2$, the rank representation of ΔZ is $(A, B, 0, 0, A, 0, A \oplus B, B)$. Strict speaking, it is $(0001, 0010, 0000, 0000, 0001, 0000, 0011, 0010)$.

Algorithm 1 *(Simple calculation method on the rank of $\mathcal{D} \cdot \Delta Y$)*

S1 *Make a rank representation of an output difference of S-function ΔZ for a given input truncated differential δX.*

S2 *Transform the rank representation of ΔZ to a rank representation of output difference of F-function ΔY by P-function.*

S3 *Hold the rank representation of ΔY that coincides with given output truncated differential δY as a candidate of ΔY. If there are several candidates, do the following checks.*
- *Discard all but one candidate, when their rank representations are the same[8].*
- *Keep the candidates whose rank is minimum.*
- *Keep the candidates whose number of the same difference patterns in the rank representation[9] is maximum, if there are several candidates whose rank is the same.*
- *Hold all the candidates that survive the above checks.*

S4 *Repeat S2 and S3 for other rank representations of ΔZ. After finishing the above process for all rank representations, all the candidates are regarded as (candidates of) ΔY.*

[8] The meaning of the same rank representations is that there is the same relations between output differences among the candidates; e.g., $(A, A, B, A \oplus B)$ and $(B, B, A \oplus B, A)$.

[9] For example, the numbers of the same difference patterns in $(A, A, B, A \oplus B, A \oplus B)$ and (A, B, B, B, A) are 2 and 3, respectively.

Table 1. Upper Bound of Bytewise Characteristic Probability of Camellia (log_2 representation)

Rounds	1	2	3	4	5	6	7	8	9	10	11
without FL/FL^{-1}-func.	0	0	-7	-21	-36	-51	-73	-93	—	—	—
with FL/FL^{-1}-func.	0	0	-7	-21	-36	-51	-51	-59	-59	-103	—

S5 *Generate an* 8×8 *matrix* $\mathcal{D}$ *by following Theorem 6, and calculate the rank of* $\mathcal{D} \cdot \Delta Y$ *for every* ΔY*.*

Step 3: Search Truncated Differential Characteristics of N Rounds

There is basically no modification of Matsui's search algorithm. The one modification is that, if FL/FL^{-1} functions are inserted, truncated differential probability is changed by following Definition 8 whenever truncated differentials pass through FL/FL^{-1} functions.

The search algorithm is terminated when (the upper bound of) truncated differentials become indistinguishable from random permutations. That is, the final round is the r-th round whose (upper bound of) maximum truncated differential probability is lower than $2^{-128+8h}$, where h is the Hamming weight of the truncated differential $(\delta X^{(r)}, \delta X^{(r+1)})$; i.e., $h = hw(\delta X^{(r)}) + hw(\delta X^{(r+1)})$. Note that, a random permutation outputs the same form of truncated differential $(\delta X^{(r)}, \delta X^{(r+1)})$ with probability $2^{-128+8h}$.

As shown in Table 1, we obtained the upper bound of best bytewise characteristic probability using the modified search algorithm. Table 1 shows that, if the number of rounds is more than 11, (reduced versions of) Camellia with FL/FL^{-1} functions is indistinguishable from random permutations in terms of truncated differential cryptanalysis. An interested thing is that Table 1 doesn't increase from the eighth round to the ninth round if FL/FL^{-1}-functions are present. This is because the input truncated differential of the ninth round becomes zero by XOR cancellation at the eighth round with probability 2^{-8}.

Thus, it is proven (from the designer's standpoint) that variants of Camellia with more than 11 rounds are secure against truncated differential cryptanalysis even if weak-key FL/FL^{-1} functions are taken into consideration. Note that, since the probability shown in Table 1 is the upper bound, we believe that Camellia is more secure than our estimation indicates.

7 Conclusion

In this paper, we studied security against truncated differential cryptanalysis from the "designer's" standpoint. Our evaluation is based on the upper bound of truncated differential probability, and we evaluated the security of Camellia against truncated differential cryptanalysis by using a modified version of Matsui's search algorithm.

We proved (from the designer's standpoint) that variants of Camellia with more than 11 rounds are secure against truncated differential cryptanalysis even if weak-key FL/FL^{-1} functions are taken into consideration. Note that, our result does NOT mean that 10-round variant of Camellia is breakable by truncated differential cryptanalysis, since the probability shown in Table 1 is the upper bound. That is, we did not show whether such effective truncated differentials exist.

References

[AIK+00] K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, and T. Tokita, "Camellia: A 128-Bit Block Cipher Suitable for Multiple Platforms," *Selected Areas in Cryptography — 7th Annual International Workshop, SAC2000*, pp.41–54 in preproceeding, 2000, (LNCS to appear).

[BS93] E. Biham and A. Shamir, "Differential Cryptanalysis of the Data Encryption Standard," Springer-Verlag, Berlin, 1993.

[ABK98] R. Anderson, E. Biham, and L. R. Knudsen, "SERPENT," *The First Advanced Encryption Standard Candidate Conference*, 1998. `http//www.cl.cam.ac.uk/~rja14/serpent.html`

[DKR97] J. Daemen, L. Knudsen, and V. Rijmen, "The block cipher SQUARE," *Fast Software Encryption — 4th International Workshop, FSE'97*, LNCS **1267**, pp.54–68, Springer-Verlag, Berlin, 1997.

[DR98] J. Daemen and V. Rijmen, "RIJNDAEL," *The First Advanced Encryption Standard Candidate Conference*, 1998. `http//www.esat. kuleuven.ac.be/~rijmen/rijndael/`

[K95] L. R. Knudesen, "Truncated and Higher Order Differentials," *Fast Software Encryption — Second International Workshop*, LNCS **1008**, pp.196–211, Springer-Verlag, Berlin, 1995.

[KB96] L. R. Knudsen and T. A. Berson, "Truncated Differentials of SAFER," *Fast Software Encryption — Third International Workshop*, LNCS **1039**, pp.15–26, Springer-Verlag, Berlin, 1996.

[KMA+00] M. Kanda, S. Moriai, K. Aoki, H. Ueda, Y. Takashima, K. Ohta, and T. Matsumoto, "*E2* – A New 128-Bit Block Cipher," *IEICE Transactions Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E83-A, No. 1, pp.48–59, 2000.

[M97] M. Matsui, "New Block Encryption Algorithm MISTY," *Fast Software Encryption — 4th International Workshop, FSE'97*, LNCS **1267**, pp.54–68, Springer-Verlag, Berlin, 1997.

[M99] M. Matsui, "Differential Path Search of the Block Cipher E2," *IEICE Technical report*, ISEC99-19, 1999-07. (In Japanese.)

[MT99] M. Matsui and T. Tokita, "Cryptanalysis of a Reduced Version of the Block Cipher E2," *Fast Software Encryption — 6th International Workshop, FSE'99*, LNCS **1636**, pp.71–80, Springer-Verlag, Berlin, 1999.

[MSA+99] S. Moriai, M. Sugita, K. Aoki, and M. Kanda, "Security of E2 against Truncated Differential Cryptanalysis," *Selected Areas in Cryptography — 6th Annual International Workshop, SAC'99*, LNCS **1758**, pp.106–117, Springer-Verlag, Berlin, 2000.

[RDP+96] V. Rijmen, J. Daemen, B. Preneel, A. Bosselaers, and E.D. Win, "The cipher SHARK," *Fast Software Encryption — Third International Workshop*, LNCS **1039**, pp.99–111, Springer-Verlag, Berlin, 1996.