Design Principles of the KASUMI Block Cipher

Johan Wallén
Helsinki University of Technology
johan.wallen@hut.fi

Abstract

In this paper, we discuss some of theory of provable security against differential and linear cryptanalysis. We also review the design principles of the block cipher KASUMI—especially its resistance against the basic forms of linear and differential cryptanalysis.

Key words. KASUMI, differential cryptanalysis, linear cryptanalysis, provable security, correlation.

1 Introduction

Since the advent of differential and linear cryptanalysis, a significant effort has been put into the design of block ciphers resistant against these powerful attacks. One branch of this research has focused on *provable* security against differential and linear cryptanalysis. The first example of a block cipher with provable security against the basic forms of differential and linear cryptanalysis under and independent round key assumption was presented by Nyberg and Knudsen [20, 18]. Later, Matsui [16] introduced a methodology for designing block ciphers with provable security against differential and linear cryptanalysis in the sence of [20, 18]. This methodology is based on the same principles as the example by Nyberg and Knudsen, but it uses some new structures, and applies the constructions recursively to the round functions to reduce the size of substitution boxes. This methodology has since been used in the design of the block ciphers MISTY [17] and KASUMI [8, 9].

In this paper, we review some of the theory of provable security against conventional differential and linear cryptanalysis under an independent round key assumption. Throughout the paper, we focus on the aspects of the theory most relevant to the design of KASUMI. We also review the design principles of KASUMI from the viewpoint of differential and linear cryptanalysis.

It is important to note that the theory discussed in this paper only gives provably security against *conventional* differential and linear cryptanalysis under an *independent subkey assumption.* The independent subkey assumption simply asserts that the round subkeys of the cipher are independent and uniformly distributed. We also make the reasonable assumption that the key is independent from the plaintext.

Although resistance against conventional differential and linear cryptanalysis does not imply that a cipher is secure against other types of attacks, it is an excellent starting point for designing block ciphers, as well as for discussing their security.

The rest of this paper is organized as follows. In Section 2 we give a self contained but brief description of conventional differential and linear cryptanalysis. In Section 3, we discuss the theory of provable resistance against differential and linear cryptanalysis most relevant to the design of KASUMI. Finally, we give a brief description of KASUMI in Section 4, and apply the results from Section 3 to it.

2 Differential and linear cryptanalysis

In this section, we give a quick overview of *differential* and *linear cryptanalysis*, and fix some terminology and notation used in the rest of the paper.

2.1 Differential cryptanalysis

Differential cryptanalysis [3] is a chosen plaintext attack that studies the propagation of input differences to output differences in iterated transformations. These difference propagations are formalized in the following definition.

Definintion 1. Let $f\colon GF(2)^n \mapsto GF(2)^m$, and let $a, a^* \in GF(2)^n$. The *difference* $a' = a + a^*$ is said to *propagate* to the difference $b' = f(a) + f(a^*)$ trough f. This is denoted by $a' \xrightarrow{f} b'$, or simply $a' \rightarrow b'$, if f is clear from context. An expression of the form $\alpha \xrightarrow{f} \beta$ is called a *differential.*

If the input difference of a pair is α, the differential $\alpha \rightarrow \beta$ can be used to predict the corresponding output difference. It is thus natural to measure the efficiency of a differential as the fraction of all inputs with difference α that results in the output difference β. Following [5], we call this fraction the *propagation ratio* of the differential.

Definintion 2. The *propagation ratio* R_p of the differential $\alpha \xrightarrow{f} \beta$ is defined by

$$R_p(\alpha \xrightarrow{f} \beta) = 2^{-n}|\{x \in GF(2)^n \mid f(x) + f(x+\alpha) = \beta\}|.$$

A generic differential attack against an r round iterated block cipher is the following.

1. Find an $r-1$ round differential $\alpha \rightarrow \beta$ with high enough propagation ratio.
2. Keep a counter for each possible round subkey k_r at round r. Initialize the counters to zero.
3. Pick a plaintext x uniformly at random and set $x^* = x + \alpha$. Encrypt the plaintexts under the unknown key K obtaining the ciphertexts y and y^*. For each possible round subkey k_r compatible with the assumed input difference β and the observed outputs y, y^* at round r, add one to the corresponding counter.

4. Repeat step 3 until some round subkeys are counted significantly more often than the others. Output these keys as the most likely candidates for the actual subkey at the last round.

There are several improvements to this basic attack that reduce the number of plaintexts needed. There are also attacks using $r - 2$ round differentials that counts on the round subkeys of the last two rounds. See [3] for further details.

Using straightforward statistical analysis, it can be shown that the correct round key can be distinguished from a randomly selected key with sufficient confidence, provided that the number of plaintexts available is inversely proportional to the propagation ratio of the differential used. Thus, a necessary condition for resistance against conventional differential attacks is that there does not exist any differential ranging over all but a few (say, 3) rounds with propagation ratio significantly larger than 2^{-n}, where n is the block size.

For key dependent functions, we consider the *average* resistance against differential cryptanalysis, and hence the average propagation ratios taken over the key set.

Definintion 3. Let $F\colon GF(2)^n \times K \mapsto GF(2)^m$ be a key dependent function. Denote $f_k(x) = F(x,k)$ for each fixed $k \in K$. Let $\alpha \in GF(2)^n, \beta \in GF(2)^m$ be constants. The *potentials* of the differentials $\alpha \xrightarrow{f_k} \beta$ and $\alpha \xrightarrow{F} \beta$ is defined by

$$DP(\alpha \xrightarrow{f_k} \beta) = R_p(\alpha \xrightarrow{f_k} \beta), \text{and}$$
$$DP(\alpha \xrightarrow{F} \beta) = \frac{1}{|K|} \sum_{k \in K} DP(\alpha \xrightarrow{f_k} \beta).$$

The *maximum potential* of F is defined by

$$DP^F_{\max} = \max_{\alpha \neq 0, \beta} DP(\alpha \xrightarrow{F} \beta).$$

Proposition 1 ([13]). *A block cipher with block length n is resistant against conventional differential attacks under an independent subkey assumption, if there does not exist any differential $\alpha \to \beta, \alpha \neq 0$ ranging over all but a few rounds, such that $DP(\alpha \to \beta) \gg 2^{-n}$.*

Note that resistance against conventional differential attacks does not imply anything about resistance against natural extensions to differential cryptanalysis, such as *impossible* [1, 2], *higher order* [12, 11] and *truncated* [11] differentials, and the *boomerang attack* [21].

2.2 Linear cryptanalysis

Linear cryptanalysis [14, 15] is a known plaintext attack that is based on "effective" linear approximate relations between the plaintext, the ciphertext, and the key. The following terminology is convenient for discussing linear cryptanalysis.

A binary *selection vector* $w \in GF(2)^n$ selects the bits i for which $w_i = 1$. The *linear combination* of the bits of a vector $x \in GF(2)^n$ selected by w can be expressed as the dot

product $w \cdot x = w_1x_1 + w_2x_2 + \cdots + w_nx_n$. For convenience, we let $l_w : GF(2)^n \mapsto GF(2)$ denote the mapping $l_w(x) = w \cdot x$. An (approximate) *linear relation* between the binary vectors $x \in GF(2)^n$ and $y \in GF(2)^m$ is an expression of the type $w \cdot x = u \cdot y$. The efficiency of the approximation is measured by its *correlation.*

Definintion 4. Let $f, g : GF(2)^n \mapsto GF(2)$ be Boolean functions. The *correlation coefficient* $c(f, g)$ of f and g is defined by

$$c(f, g) = 2^{-n}(|\{x \in GF(2)^n \mid f(x) = g(x)\}| - |\{x \in GF(2)^n \mid f(x) \neq g(x)\}|).$$

In the basic form of linear cryptanalysis of an r round iterated block cipher, the analyst tries to find a linear approximation over $r - 2$ rounds from the second round to the second to last round—that is, an approximation of the form

$$a \cdot X + b \cdot Y + c \cdot k = 0, \tag{1}$$

where $X = f_1(x, k_1)$, $y = f_r(Y, k_r)$ is the ciphertext, and $k = (k_2, \ldots, k_{r-1})$ is a vector of all the unknown round keys used at rounds 2 to $r - 1$. Given N known plaintext, the parts of the round keys k_1 and k_r relevant to the approximation can be found by trying all possible round subkeys at rounds 1 and r, and counting the number N_0 of plaintext for which

$$a \cdot f_1(x, k_1) + b \cdot f_r^{-1}(y, k_r) = 0. \tag{2}$$

holds. The round subkeys that maximizes $|N_0/N - 1/2|$ are chosen as the most likely candidates.

In [14, 18], if was shown that the number of known plaintexts needed for the attack above is inversely proportional to the average taken over k of the square of the correlations between $a \cdot X$ and $b \cdot Y$. In [18], it was shown that this average equals the sum taken over c of the squares of the correlations of (1). It is thus natural to consider the family of all linear approximations of (1), when c ranges over all possible values. This family of approximations (without any reference to the selection vector c) is called a *linear hull* [18], and is denoted by $b \longleftarrow a$. Its *potential* is defined to be the above mentioned average. Thus, a block cipher is secure against conventional linear cryptanalysis, if there there does not exist any linear hull ranging over all but a few rounds, such that its potential is significantly larger that 2^{-n}, where n is the block size.

Definintion 5. Let $F : GF(2)^n \times K \mapsto GF(2)^m$ be a key dependent function. Denote $f_k(x) = F(x, k)$ for each fixed $k \in K$. Let $w \in GF(2)^n, u \in GF(2)^m$ be constants. The *potential* of the *linear approximations* $u \xleftarrow{f_k} w$ and $u \xleftarrow{F} w$ is defined by

$$LP(u \xleftarrow{f_k} w) = |c(l_u \circ f_k, l_w)|^2, \quad \text{and}$$
$$LP(u \xleftarrow{F} w) = \frac{1}{|K|} \sum_{k \in K} LP(u \xleftarrow{f_k} w).$$

When F is clear from context, $u \xleftarrow{F} w$ is written as $u \longleftarrow w$. The potential of the best linear approximation of F is given by

$$LP^F_{\max} = \max_{u \neq 0, w} LP(u \xleftarrow{F} w).$$

Proposition 2 ([18]). *A block cipher with block length n is resistant against conventional linear cryptanalysis under an independent subkey assumption, if there does not exist any linear approximation $u \longleftarrow w, u \neq 0$ ranging over all but a few rounds, such that $LP(u \longleftarrow w) \gg 2^{-n}$.*

3 Block ciphers with provable security against conventional differential and linear attacks

Since the advent of differential and linear cryptanalysis, several researchers have proposed different design strategies to achieve resistance against differential and linear cryptanalysis. In [20], a DES-like block cipher prototype with provably low differential potentials, and thus with provable security against conventional differential attacks under an independent round key assumption, was presented. In [18], it was shown that this cipher also has provably low linear potentials, and thus also is provably secure against conventional linear attacks under the same assumptions.

In [16] some other constructions for block ciphers with provable security against conventional differential and linear attacks was presented. One of the main contributions of [16] was to use round functions with recursive structure, which reduces the size of substitution boxes, and thus makes the cipher more suitable for software and hardware implementation without loosing any of the provable properties of the cipher in [20].

3.1 Preliminaries

Before we derive the linear and differential potentials of the constructions used in KASUMI, we review some useful results. The results in this section have been extensively discussed in the literature; see [18, 4, 6, 5, 19]. For completeness, we give direct proofs of all non-trivial results needed.

The routine verification of the following three lemmata is left to the reader.

Lemma 1. *Let $f\colon GF(2)^n \mapsto GF(2)^m$. For all $u \in GF(2)^m, w \in GF(2)^n$,*

$$c(l_u \circ f, l_w) = 2^{-n} \sum_{x \in GF(2)^n} (-1)^{u \cdot f(x) + w \cdot x}.$$

Lemma 2. *Let $F\colon GF(2)^n \times K \mapsto GF(2)^n$ be bijective for all fixed $k \in K$. Then*

$$\sum_{w \in GF(2)^n} LP(u \xleftarrow{F} w) = 1, \qquad \sum_{\beta \in GF(2)^n} DP(\alpha \xrightarrow{F} \beta) = 1,$$

$$\sum_{u \in GF(2)^n} LP(u \xleftarrow{F} w) = 1, \text{ and} \qquad \sum_{\alpha \in GF(2)^n} DP(\alpha \xrightarrow{F} \beta) = 1.$$

Lemma 3. *Let $f\colon GF(2)^n \mapsto GF(2)^m$. For all $u \in GF(2)^m$ and $x \in GF(2)^n$,*

$$(-1)^{u \cdot f(x)} = \sum_{w \in GF(2)^n} c(l_u \circ f, l_w)(-1)^{w \cdot x}.$$

Note that Lemma 3 asserts that a Boolean function is uniquely determined by its set of correlation coefficients to all linear functions.

Lemma 4. *Let* $f\colon GF(2)^n \mapsto GF(2)^p$ *and* $g\colon GF(2)^p \mapsto GF(2)^m$. *Then*

$$c(l_u \circ (g \circ f), l_w) = \sum_{v \in GF(2)^p} c(l_u \circ g, l_v) c(l_v \circ f, l_w).$$

Proof. Lemma 3 gives

$$\begin{aligned}(-1)^{u \cdot g(f(x))} &= \sum_{v \in GF(2)^p} c(l_u g, l_v)(-1)^{v \cdot f(x)} \\ &= \sum_{v \in GF(2)^p} c(l_u g, l_v) \sum_{w \in GF(2)^n} c(l_v f, l_w)(-1)^{w \cdot x} \\ &= \sum_{w \in GF(2)^n} \left(\sum_{v \in GF(2)^p} c(l_u g, l_v) c(l_v f, l_w) \right) (-1)^{w \cdot x}.\end{aligned}$$

□

An immediate consequence of Lemma 4 is

Lemma 5. *Let* $f\colon GF(2)^n \mapsto GF(2)^m$ *be a Boolean function, and let* $a \in GF(2)^n, b \in GF(2)^m$ *be constants. Define* $f_a(x) = f(x + a)$ *and* $f^b(x) = f(x) + b$. *Then*

$$\begin{aligned}c(l_u \circ f_a, l_w) &= (-1)^{w \cdot a} c(l_u \circ f, l_w), \text{ and} \\ c(l_u \circ f^b, l_w) &= (-1)^{u \cdot b} c(l_u \circ f, l_w).\end{aligned}$$

The simple proof of the following lemma is left to the reader.

Lemma 6. *Let* $f, g\colon GF(2)^n \mapsto GF(2)^m$. *Then*

$$2^{-n} \sum_{x \in GF(2)^n} (-1)^{u \cdot f(x) + u \cdot g(x)} = \sum_{w \in GF(2)^n} c(l_u \circ f, l_w) c(l_u \circ g, l_w).$$

There is a remarkable correspondence between the propagation ratio and the linear correlations of a Boolean mapping, namely

Lemma 7. *Let* $f\colon GF(2)^n \mapsto GF(2)^m$. *For all* $\alpha \in GF(2)^n, \beta \in GF(2)^m$,

$$R_p(\alpha \xrightarrow{f} \beta) = 2^{-m} \sum_{u \in GF(2)^m, w \in GF(2)^n} (-1)^{w \cdot \alpha + u \cdot \beta} c(l_u \circ f, l_w)^2.$$

Proof. Note that $2^{-m}\sum_{u\in GF(2)^m}(-1)^{u\cdot(f(x)+f(x+\alpha)+\beta)} = 1$ if $f(x)+f(x+\alpha)+\beta = 0$, and 0 otherwise. Let δ denote the function $\delta(0) = 1$, and $\delta(x) = 0$ for $x \neq 0$. We have

$$\begin{aligned} R_p(\alpha \xrightarrow{f} \beta) &= 2^{-n} \sum_{x\in GF(2)^n} \delta(f(x) + f(x+\alpha) + \beta) \\ &= 2^{-n} \sum_x 2^{-m} \sum_{u\in GF(2)^m} (-1)^{u\cdot(f(x)+f(x+\alpha)+\beta)} \\ &= 2^{-m} \sum_u (-1)^{u\cdot\beta} 2^{-n} \sum_x (-1)^{u\cdot(f(x)+f(x+\alpha))}. \end{aligned}$$

Denote $f_\alpha(x) = f(x+\alpha)$. Lemma 6 then gives

$$\begin{aligned} R_p(\alpha \xrightarrow{f} \beta) &= 2^{-m} \sum_u (-1)^{u\cdot\beta} \sum_{w\in GF(2)^n} c(l_u f, l_w) c(l_u f_\alpha, l_w) \\ &= 2^{-m} \sum_{u,w} (-1)^{w\cdot\alpha + u\cdot\beta} c(l_u f, l_w)^2. \end{aligned}$$

□

Given Lemma 7, the following result is obvious.

Lemma 8. *Let $F\colon GF(2)^n \times K \mapsto GF(2)^m$ be a key dependent function. Then*

$$DP(\alpha \xrightarrow{F} \beta) = 2^{-m} \sum_{u\in GF(2)^m, w\in GF(2)^n} (-1)^{w\cdot\alpha+u\cdot\beta} LP(u \xleftarrow{F} w).$$

3.2 Some special mappings

Using the results from the previous subsection, we can easily determine the maximum differential and linear potentials of some mappings important to the design of KASUMI. We start with considering the composition of key dependent mappings.

Lemma 9. *Let $F\colon GF(2)^n \times GF(2)^n \mapsto GF(2)^n$ and $G\colon GF(2)^n \times GF(2)^n \mapsto GF(2)^n$ be key dependent functions of the form $F(x,k) = f(x+k), G(x,k) = g(x+k)$, where $f, g\colon GF(2)^n \mapsto GF(2)^n$ are bijections. Then*

$$LP(u \xleftarrow{G\circ F} w) = \sum_{v\in GF(2)^n} LP(u \xleftarrow{f} v) LP(v \xleftarrow{g} w).$$

Proof. Denote $F_k = f(x+k)$ and $G_k = g(x+k)$. Note that $c(l_u \circ (G_{k_2} \circ F_{k_1}), l_w) = \sum_v (-1)^{v\cdot k_2} c(l_u \circ g, l_v) c(l_v \circ F_{k_1}, l_w)$. Thus

$$\begin{aligned} &2^{-2n} \sum_{k_1,k_2} c(l_u \circ (G_{k_2} \circ F_{k_1}), l_w)^2 = \\ &\quad = 2^{-2n} \sum_{k_1} \sum_{v,v'} c(l_u \circ g, l_v) c(l_u \circ g, l_{v'}) c(l_v \circ F_{k_1}, l_w) c(l_{v'} \circ F_{k_1}, l_w) \sum_{k_2} (-1)^{k_2\cdot(v+v')} \\ &\quad = 2^{-n} \sum_{k_1} \sum_v c(l_u \circ g, l_v)^2 c(l_v \circ F_{k_1}, l_w)^2 \end{aligned}$$

Since $c(l_v \circ F_{k_1}, l_w) = (-1)^{w \cdot k_1} c(l_v \circ f, l_w)$, this equals

$$2^{-n} \sum_{k_1} \sum_{v} c(l_u \circ g, l_w)^2 c(l_v \circ f, l_w)^2 = \sum_{v} c(l_u \circ g, l_w)^2 c(l_v \circ f, l_w)^2.$$

□

An immediate consequence of Lemma 9 is

Theorem 1 ([18]). *Let $F\colon GF(2)^n \times GF(2)^n \times K_1$ and $G\colon GF(2)^n \times GF(2)^n \times K_2$ be key dependent functions of the type $F(x, k, k') = f(x + k, k'), G(x, k, k') = g(x + k, k')$, where $f\colon GF(2)^n \times K_1 \mapsto GF(2)^n$ and $g\colon GF(2)^n \times K_2 \mapsto GF(2)^n$ are bijective for all fixed $k_1 \in K_1, k_2 \in K_2$. Then*

$$LP(u \xleftarrow{G \circ F} w) = \sum_{v \in GF(2)^n} LP(u \xleftarrow{g} v) LP(v \xleftarrow{f} w)$$

A straightforward application of Lemma 8 now gives

Corollary 1 ([13]). *Let F and G be as in Theorem 1. Then*

$$DP(\alpha \xrightarrow{G \circ F} \beta) = \sum_{\xi \in GF(2)^n} DP(\alpha \xrightarrow{f} \xi) DP(\xi \xrightarrow{g} \beta).$$

The following corollary is a formalization of the folk theorem that increasing the number of rounds does not weaken a cipher.

Corollary 2. *Let F and G be as in Theorem 1. Then*

$$LP^{G \circ F}_{\max} \leq \min(LP^f_{\max}, LP^g_{\max}), \text{ and}$$
$$DP^{G \circ F}_{\max} \leq \min(DP^f_{\max}, DP^g_{\max}).$$

Proof. We give a proof of the first inequality. The proof of the second inequality is similar, and left to the reader. Assume, without loss of generality, that $LP^f_{\max} \leq LP^g_{\max}$. Since g is bijective, $LP(u \xleftarrow{g} 0) = 1$, if $u = 0$, and 0 otherwise. Thus, Theorem 1 gives for all $u \neq 0$

$$\begin{aligned} LP(u \xleftarrow{G \circ F} w) &= \sum_{v \in GF(2)^n} LP(u \xleftarrow{g} v) LP(v \xleftarrow{f} w) \\ &= \sum_{v \in GF(2)^n - \{0\}} LP(u \xleftarrow{g} v) LP(v \xleftarrow{f} w) \\ &\leq LP^f_{\max} \sum_{v \in GF(2)^n - \{0\}} LP(u \xleftarrow{g} v) \leq LP^f_{\max}. \end{aligned}$$

□

Next, we turn our attention to the function in Figure 1.

Theorem 2 ([16]). *Let F be the r round function in Figure 1 with $r \geq 3$, where each $F_i\colon GF(2)^n \times GF(2)^n \times K_i \mapsto GF(2)^n$ is of the form $F_i(x, k_i, k_i') = f_i(x + k_i, k_i')$, and each $f_i\colon GF(2)^n \times K_i \mapsto GF(2)^n$ is bijective for all fixed $k_i' \in K_i$. If $LP^{f_i}_{\max} \leq p$ (respective $DP^{f_i}_{\max} \leq p$) for each i, then $LP^{F}_{\max} \leq p^2$ (respective $DP^{F}_{\max} \leq p^2$).*

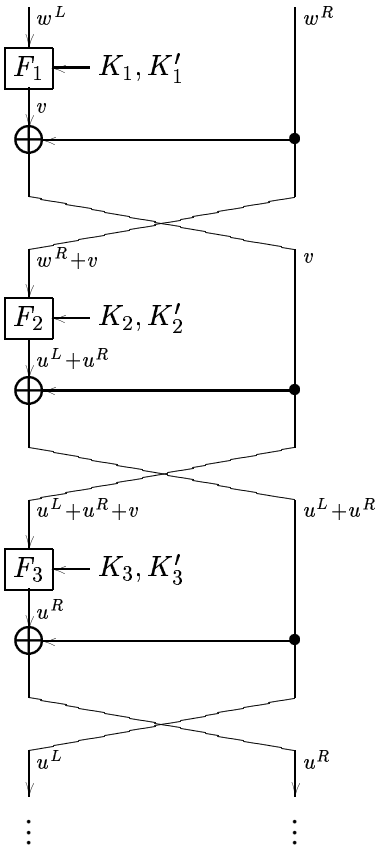


Figure 1: The function in Theorem 2.

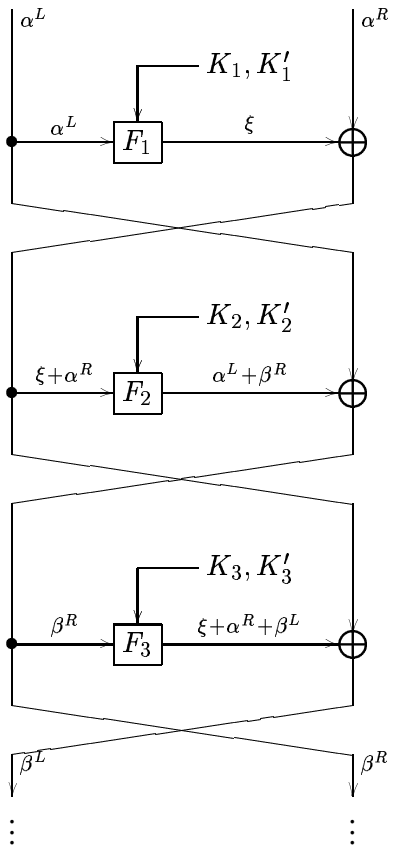


Figure 2: The function in Theorem 3.

Proof. We give a proof of the linear case; the proof of the differential case is similar, and can be found in [16].

We first consider the case $r = 3$. The general case then follows from Corollary 2.

Denote by x^L and x^R the left and right halves of x, respectively. Let the output selection vector of F be $u = (u^L, u^R) \neq 0$, and the input selection vector be $w = (w^L, w^R) \neq 0$. If the output selection vector of F_1 is v, it can easily be seen that we have the individual round approximations $v \xleftarrow{F_1} w^L$, $u^L + u^R \xleftarrow{F_2} w^R + v$, and $u^R \xleftarrow{F_3} u^L + u^R + v$. Theorem 1 gives

$$LP(u \xleftarrow{F} w) = \sum_{v \in GF(2)^n} LP(v \xleftarrow{f_1} w^L) LP(u^L + u^R \xleftarrow{f_2} w^R + v) LP(u^R \xleftarrow{f_3} u^L + u^R + v).$$

There are three (disjoint) special cases to consider: $v = 0$, $u^L + u^R = 0$, and $u^R = 0$.

1. If $v = 0$, $w^L = 0$, and thus $w^R \neq 0$. Since $w^R + v \neq 0$, $u^L + u^R \neq 0$. Since $u^L + u^R + v \neq 0$, $u^R \neq 0$ too. We have

$$\begin{aligned} LP(u \xleftarrow{F} w) &= LP(u^L + u^R \xleftarrow{f_2} w^R) LP(u^R \xleftarrow{f_3} u^L + u^R) \\ &\leq LP^{f_2}_{\max} LP^{f_3}_{\max} \leq p^2. \end{aligned}$$

2. If $u^L + u^R = 0$, $u^L = u^R \neq 0$, and $u^L + u^R + v = v$. Since $w^R + v = 0$, $v \neq 0$. ($v = 0$ would give $w^L = w^R = 0$). We have

$$\begin{aligned} LP(u \xleftarrow{F} w) &= LP(v \xleftarrow{f_1} w^L) LP(u^R \xleftarrow{f_3} v) \\ &\leq LP^{f_1}_{\max} LP^{f_3}_{\max} \leq p^2. \end{aligned}$$

3. If $u^R = 0$, $u^L \neq 0$. Thus $u^L + u^R = u^L \neq 0$. Since $u^L + u^R + v = 0$, $v = u^L + u^R = u^L \neq 0$, and we have

$$\begin{aligned} LP(u \xleftarrow{F} w) &= LP(v \xleftarrow{f_1} w^L) LP(u^L \xleftarrow{f_2} w^R + v) \\ &\leq LP^{f_1}_{\max} LP^{f_2}_{\max} \leq p^2. \end{aligned}$$

4. Otherwise, v, $u^L + u^R$, and u^R are all non-zero, and

$$\begin{aligned} LP(w \xleftarrow{F} u) &\leq \sum_{v \in GF(2)^n} LP(v \xleftarrow{f_1} w^L) LP^{f_2}_{\max} LP^{f_3}_{\max} \\ &= LP^{f_2}_{\max} LP^{f_3}_{\max} \leq p^2. \end{aligned}$$

Note that cases 1–3 indeed are disjoint. □

Finally, we consider the Feistel cipher in Figure 2.

Theorem 3 ([20, 18]). *Let F be the r round ($r \geq 3$) Feistel cipher in Figure 2, where each $F_i \colon GF(2)^n \times GF(2)^n \times K_i \mapsto GF(2)^n$ is of the form $F_i(x, k_i, k_i') = f_i(x + k_i, k_i')$, and each $f_i \colon GF(2)^n \times K_i \mapsto GF(2)^n$ is bijective for all fixed $k_i' \in K_i$. If $LP^{f_i}_{\max} \leq p$ (respective $DP^{f_i}_{\max} \leq p$) for each i, then $LP^F_{\max} \leq p^2$ (respective $DP^F_{\max} \leq p^2$).*

Proof. We give a proof of the differential case; the proof of the linear case is similar to the proof of Theorem 2.

We first concider the case $r = 3$. Let the input difference of F be $\alpha = (\alpha^L, \alpha^R) \neq 0$ and the output difference be $\beta = (\beta^L, \beta^R) \neq 0$. If the output difference of F_1 is ξ, it is easy to see that that we have the individual round differentials $\alpha^L \xrightarrow{F_1} \xi$, $\xi + \alpha^R \xrightarrow{F_2} \alpha^L + \beta^R$, and $\beta^R \xrightarrow{F_3} \xi + \alpha^R + \beta^L$. Corollary 1 gives

$$\begin{aligned} &DP(\alpha \xrightarrow{F} \beta) = \\ &\quad \sum_{\xi \in GF(2)^n} DP(\alpha^L \xrightarrow{f_1} \xi) DP(\xi + \alpha^R \xrightarrow{f_2} \alpha^L + \beta^R) DP(\beta^R \xrightarrow{f_3} \xi + \alpha^R + \beta^L). \end{aligned}$$

There are three (disjoint) special cases to consider: $\alpha^L = 0$, $\xi + \alpha^R = 0$, and $\beta^R = 0$.

1. If $\alpha^L = 0$, $\alpha^R \neq 0$ and $\xi = 0$. The input difference for f_2 is $\xi + \alpha^R = \alpha^R \neq 0$. Since f_2 is bijective, $\alpha^L + \beta^R = \beta^R \neq 0$. Thus, the input difference for f_3 is also non-zero. We have

$$\begin{aligned} DP(\alpha \xrightarrow{F} \beta) &= DP(\alpha^R \xrightarrow{f_2} \beta^R) DP(\beta^R \xrightarrow{f_3} \alpha^R + \beta^L) \\ &\leq DP^{f_2}_{\max} DP^{f_3}_{\max} \leq p^2. \end{aligned}$$

2. If $\xi + \alpha^R = 0$, $\xi = \alpha^R \neq 0$ (since f_1 is bijective, $\xi = 0$ would give $\alpha^L = 0$). Thus, $\alpha^L \neq 0$. We also have that $\alpha^L + \beta^R = 0$. Thus, $\beta^R = \alpha^L \neq 0$. We have

$$\begin{aligned} DP(\alpha \xrightarrow{F} \beta) &= DP(\alpha^L \xrightarrow{f_1} \alpha^R) DP(\beta^R \xrightarrow{f_3} \beta^L) \\ &\leq DP^{f_1}_{\max} DP^{f_3}_{\max} \leq p^2. \end{aligned}$$

3. If $\beta^R = 0$, $\beta^L \neq 0$ and $\xi + \alpha^R + \beta^L = 0$. Thus, $\xi + \alpha^R = \beta^L \neq 0$. Now $\alpha^L + \beta^R = \alpha^L \neq 0$ ($\alpha^L = 0$ would imply that $\xi + \alpha^R = 0$). Since $\alpha^L \neq 0$, $\xi = \alpha^R + \beta^L \neq 0$ too, and we have

$$\begin{aligned} DP(\alpha \xrightarrow{F} \beta) &= DP(\alpha^L \xrightarrow{f_1} \alpha^R + \beta^L) DP(\beta^L \xrightarrow{f_2} \alpha^L) \\ &\leq DP^{f_1}_{\max} DP^{f_2}_{\max} \leq p^2. \end{aligned}$$

4. Otherwise, $\alpha^L \neq 0$, $\xi + \alpha^R \neq 0$, and $\beta^R \neq 0$. We have

$$\begin{aligned} DP(\alpha \xrightarrow{F} \beta) &\leq \sum_{\xi} DP(\alpha^L \xrightarrow{f_1} \xi) DP^{f_2}_{\max} DP^{f_3}_{\max} \\ &= DP^{f_2}_{\max} DP^{f_3}_{\max} \leq p^2. \end{aligned}$$

The case $r > 3$ now follows from Corollary 2. □

Note that the constructions in this subsection can be applied recursively. For example, if the round functions in Figure 2 are of the type in Figure 1, and all their subfunctions have linear (or differential) potentials bounded by p, Theorems 2 and 3 immediately gives that the linear (or differential) potentials are bounded by p^4. This construction can be repeated until the subfunctions become small enough to be implemented as substitution boxes. An upper bound for the linear and differential potentials of the whole cipher can easily be determined from the corresponding properties of the smallest subfunctions. These kinds of constructions, originally suggested in [16], are precisely the constructions used in MISTY and KASUMI.

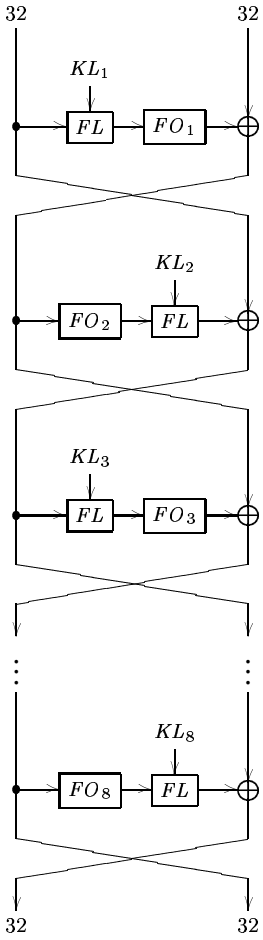
4 Description of KASUMI

KASUMI [8] is an 8 round Feistel network with a block size of 64 bits, and a key length of 128 bits that faithfully follows the constructions in Theorems 2 and 3. The encryption function is shown in Figure 3.

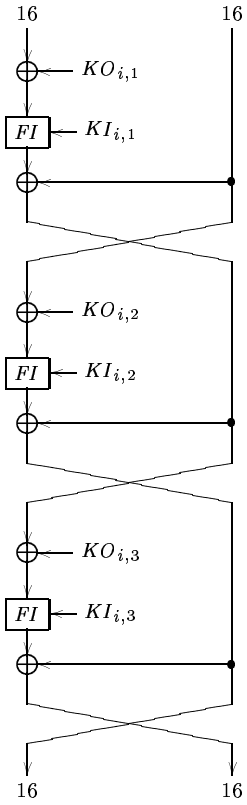
The round function consists of an outer function FO similar to the function in Theorem 2 (Figure 3(b)). This outer function is in turn built from an inner function FI (Figure 3(c)).

The FI function has a quite uncommon structure: it spits its input into two parts of unequal sizes, and uses two substitution boxes of unequal sizes. This makes algebraic analysis such as the *interpolation attack* [10] harder.

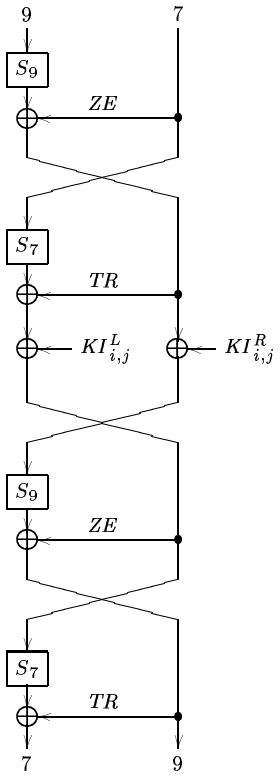
The round subkeys are derived from the 128 bit key using a straightforward affine key schedule.



(a) The encryption function.



(b) The FO function.



(c) The FI function.

Figure 3: The KASUMI encryption function. FL: $GF(2)^{32} \times GF(2)^{32} \mapsto GF(2)^{32}$ is bijective and linear for each KL. S_9 and S_7 are bijective substitution boxes over $GF(2)^9$ and $GF(2)^7$, respectively. ZE extends a 7 bit word to a 9 bit word by adding two zero bits to the most significant end. TR truncates a 9 bit word to a 7 bit word by discarding the two most significant bits.

From the tables for the S boxes S_7 and S_9 given in [8], it is easy to compute that $LP^{S_7}_{\max} = DP^{S_7}_{\max} = 2^{-6}$, and $LP^{S_9}_{\max} = DP^{S_9}_{\max} = 2^{-8}$. The simple structure of the FI function makes it possible to derive tight upper bounds for $DP^{FI}_{\max}$ and $LP^{FI}_{\max}$, much in the same way as in the proof of Theorem 2.

If the non-zero input difference to FI is $\alpha = (\alpha^L, \alpha^R) \in GF(2)^9 \times GF(2)^7$, and the

output difference is $\beta = (\beta^L, \beta^R) \in GF(2)^7 \times GF(2)^9$, it is easy to see that

$$DP(\alpha \xrightarrow{FI} \beta) = \sum_{\xi^L, \xi^R} DP(\alpha_L \xrightarrow{S_9} \xi^L + ZE(\alpha^R)) DP(\alpha^R \xrightarrow{S_7} \xi^R + TR(\xi^L)) \cdot$$
$$DP(\xi^L \xrightarrow{S_9} \beta^R + ZE(\xi^R)) DP(\xi^R \xrightarrow{S_7} \beta^L + TR(\beta^R)).$$

There are three cases to consider. If $\alpha^L = 0$, $\alpha^R \neq 0$, and $\xi^L = ZE(\alpha^R) \neq 0$. We have

$$\begin{aligned} DP(\alpha \xrightarrow{FI} \beta) &\leq DP^{S_7}_{\max} DP^{S_9}_{\max} \sum_{\xi^R} DP(\xi^R \xrightarrow{S_7} \beta^L + TR(\beta^R)) \\ &= DP^{S_7}_{\max} DP^{S_9}_{\max} = 2^{-14}. \end{aligned}$$

If $\alpha^R = 0$, $\alpha^L \neq 0$, and $\xi^L \neq 0$. Thus, $\xi^R = TR(\xi^L)$, and

$$\begin{aligned} DP(\alpha \xrightarrow{FI} \beta) &\leq DP^{S_9}_{\max} DP^{S_9}_{\max} \sum_{\xi^L} DP(TR(\xi^L) \xrightarrow{S_7} \beta^L + TR(\beta^R)) \\ &= DP^{S_9}_{\max} DP^{S_9}_{\max} \cdot 4 = 2^{-14}. \end{aligned}$$

Otherwise, α^L and α^R are both non-zero, and

$$\begin{aligned} &DP(\alpha \xrightarrow{FI} \beta) \\ &\quad \leq DP^{S_9}_{\max} DP^{S_7}_{\max} \sum_{\xi^R} DP(\xi^R \xrightarrow{S_7} \beta^L + TR(\beta^R)) \sum_{\xi^L} DP(\xi^L \xrightarrow{S_9} \beta^R + ZE(\xi^R)) \\ &\quad = DP^{S_9}_{\max} DP^{S_7}_{\max} = 2^{-14}. \end{aligned}$$

We conclude that $DP^{FI}_{\max} \leq 2^{-14}$.

Similarly, it can be seen that

$$\begin{aligned} &LP(u \xleftarrow{FI} w) = \\ &\quad \sum_{v^L, v^R} LP(u^L \xleftarrow{S_7} v^R + TR(u^R) + u^L) LP(u^R + ZE(u^L) \xleftarrow{S_9} v^L) \cdot \\ &\qquad LP(v^R \xleftarrow{S_7} w^R + TR(v^L) + v^R) LP(v^L + ZE(v^R) \xleftarrow{S_9} w^L). \end{aligned}$$

This expression can also be bounded from above by 2^{-14} for all non-zero u. Thus, $LP^{FI}_{\max} \leq 2^{-14}$.

Theorems 2 and 3 now immediately gives the estimates $DP^{\text{KASUMI}}_{\max} \leq (2^{-14})^4 = 2^{-56}$ and $LP^{\text{KASUMI}}_{\max} \leq (2^{-14})^4 = 2^{-56}$.

5 Conclusions

The theory of provable security discussed in this paper is very elegant, and provides a useful framework for block cipher design. The approach taken here can, however, be criticized.

First of all, the independent subkey assumption is in any realistic block cipher only a useful approximation. Second, we have only considered the *average* differential and linear potentials. There might still be *weak keys* for which the propagation ratios and correlations are higher. Third, we have only considered *conventional* linear and differential cryptanalysis. This does not imply anything about resistance against other attacks.

On the other hand, it should be remembered that the bounds in Theorem 3 holds already after three rounds—the effect of rest of the rounds is more or less ignored. We know that the additional rounds does not weaken the cipher, but we have not claimed that they strengthen it either. It is, however, very reasonable to assume that these additional rounds indeed increase the resistance against linear and differential cryptanalysis.

With these reservations, we believe that the theory discussed in this paper gives a deeper understanding of the security of block ciphers and provides a sound foundation for block cipher design.

We finally note that very similar design principles have been used in the Rijndael block cipher [7] to achieve resistance against differential and linear cryptanalyis.

Acknowledgments

I would like to thank Kaisa Nyberg for valuable comments on the draft of this paper, as well as for pointing out an error in a previous version of the proof of Theorem 1.

References

[1] Eli Biham, Alex Biryukov, and Adi Shamir. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In *Advances in Cryptology—Eurocrypt '99*, volume 1592 of LNCS, pages 12–23. Springer-Verlag, 1999.

[2] Eli Biham, Alex Biryukov, and Adi Shamir. Miss in the middle attacks on IDEA and Khufu. In *Fast Software Encryption '99*, volume 1636 of LNCS, pages 124–138. Springer-Verlag, 1999.

[3] Eli Biham and Adi Shamir. *Differential Cryptanalysis of the Data Encryption Standard*. Springer-Verlag, 1993.

[4] Florent Chabaud and Serge Vaudenay. Links between differential and linear cryptanalysis. In *Advances in Cryptology—Eurocrypt '94*, volume 950 of LNCS, pages 356–365. Springer-Verlag, 1995.

[5] Joan Daemen. Propagation and correlation. In *Cipher and Hash Function Design. Strategies Based on Linear and Differential Cryptanalysis*, chapter 5. Katholieke Universiteit Leuven, March 1995. Available from the Rijndael page, `http://www.esat.kuleuven.ac.be/~rijmen/rijndael/`.

[6] Joan Daemen, René Govaerts, and Joos Vandewalle. Correlation matrices. In *Fast Software Encryption '94*, volume 1008 of LNCS, pages 275–285. Springer-Verlag, 1995.

[7] Joan Daemen and Vincent Rijmen. The Rijndael block cipher, 1999. Available from `http://www.esat.kuleuven.ac.be/~rijmen/rijndael/`.

[8] ETSI/SAGE. KASUMI specification. Specification of the 3GPP Confidentiality and Integrity Algorithms Document 2, ETSI/SAGE, December 1999. Available from `http://www.etsi.org/dvbandca/3GPP/3gppspecs.htm`.

[9] ETSI/SAGE. f_8 and f_9 specification. Specification of the 3GPP Confidentiality and Integrity Algorithms Document 1, ETSI/SAGE, September 2000. Available from `http://www.etsi.org/dvbandca/3GPP/3gppspecs.htm`.

[10] Thomas Jakobsen and Lars Knudsen. The interpolation attack on block ciphers. In *Fast Software Encryption '97*, volume 1267 of LNCS, pages 28–40. Springer-Verlag, 1997.

[11] Lars Knudsen. Truncated and higher order differentials. In *Fast Software Encryption '95*, volume 1008 of LNCS, pages 196–210. Springer-Verlag, 1995.

[12] Xuejia Lai. Higer order derivatives and differential cryptanalysis. In *Communication and Cryptography*, pages 227–233. Kluwer Academic Publishers, 1994.

[13] Xuejia Lai, James Massey, and Sean Murphy. Markov ciphers and differential cryptanalysis. In *Advances in Cryptology—Eurocrypt '91*, volume 547 of LNCS, pages 17–38. Springer-Verlag, 1991.

[14] Mitsuru Matsui. Linear cryptanalysis method for DES cipher. In *Advances in Cryptology—Eurocrypt '93*, volume 765 of LNCS, pages 386–397. Springer-Verlag, 1993.

[15] Mitsuru Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology—Crypto '94*, volume 839 of LNCS, pages 1–11. Springer-Verlag, 1994.

[16] Mitsuru Matsui. New structure of block ciphers with provable security against differential and linear cryptanalysis. In *Fast Software Encryption '96*, volume 1039 of LNCS, pages 205–218. Springer-Verlag, 1996.

[17] Mitsuru Matsui. New block encyption algorithm MISTY. In *Fast Software Encryption '97*, volume 1267 of LNCS, pages 54–68. Springer-Verlag, 1997.

[18] Kaisa Nyberg. Linear approximation of block ciphers. In *Advances in Cryptology—Eurocrypt '94*, volume 950 of LNCS, pages 439–444. Springer-Verlag, 1995.

[19] Kaisa Nyberg. Correlation theorems in cryptanalysis. *Discrete Applied Mathematics*, 2000. To appear.

[20] Kaisa Nyberg and Lars Knudsen. Provable security against a differental attack. *Journal of Cryptology*, 8(1):27–37, 1995.

[21] David Wagner. The boomerang attack. In *Fast Software Encryption '99*, volume 1636 of LNCS, pages 156–170. Springer-Verlag, 1999.