$$\begin{tikzcd} & \mathbb{F}_2^n \arrow[d, "\oplus"'] & & & k \arrow[d, mapsto] & & \\ \mathbb{F}_2^n \arrow[r] & \mathbb{F}_2^n \arrow[r, "f"] & \mathbb{F}_2^n & a \arrow[r, mapsto] & b \arrow[r, mapsto] & \boxed{S} \arrow[r] & c \end{tikzcd}$$

Figure 5.2: A (much too) simple example—The graphics here and later represent the map f sometimes by the S-box S in the elementwise assignments.

5.2 Example A: A One-Round-Cipher

We consider examples that are much too simple for real world applications but illustrate the principles of linear cryptanalysis in an easily intelligible way. We always assume round functions of the type $f(a+k)$, that is we add the key—or an n-bit part of it—to the plaintext before applying a bijective S-box $f\colon \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n$. This is a quite special method of bringing the key into play but nevertheless realistic. The paradigmatic sample ciphers LUCIFER, DES, and AES do so, the term used with AES [1] is "key-alternating cipher structure".

The simplest model is encryption by the formula

$$c = f(a+k),$$

see Figure 5.2. This example is pointless because one block of known plaintext gives a solution for k:

$$k = f^{-1}(c) + a.$$

Note that the attacker knows the inverse map f^{-1} that is part of the decryption algorithm. (One-way encryption methods that assume that f^{-1} is not efficiently deducible from f are the subject of another part of cryptography, see Part III, Chapter 6, of these lecture notes.)

The somewhat more involved example A stops this attack:

$$c = f(a + k^{(0)}) + k^{(1)}$$

(see Figure 5.3). This is the simplest example for which the method of linear cryptanalysis makes sense: Let (α, β) be a pair of linear forms with

$$\beta \circ f(x) \overset{p}{\approx} \alpha(x), \qquad (3)$$

where the symbol $\overset{p}{\approx}$ reads as "equal with probability p", or in other words

$$p = p_{f,\alpha,\beta} := \frac{1}{2^n} \cdot \#\{x \in \mathbb{F}_2^n \mid \beta \circ f(x) = \alpha(x)\}.$$

$$\begin{tikzcd}
 & \mathbb{F}_2^n \arrow[d, "\bigoplus"'] & & \mathbb{F}_2^n \arrow[d, "\bigoplus"'] \\
\mathbb{F}_2^n \arrow[r] & \mathbb{F}_2^n \arrow[r, "f"] & \mathbb{F}_2^n \arrow[r] & \mathbb{F}_2^n \\
 & k^{(0)} \arrow[d, mapsto] & & k^{(1)} \arrow[d, mapsto] \\
a \arrow[r, mapsto] & b \arrow[r, mapsto, "\boxed{\mathrm{S}}"] & b' \arrow[r, mapsto] & c
\end{tikzcd}$$

Figure 5.3: Example A

$$\begin{tikzcd}
\mathbb{F}_2^n \arrow[rr, "f"] \arrow[dr, "\alpha"'] & & \mathbb{F}_2^n \arrow[dl, "\beta"] \\
 & \mathbb{F}_2 \arrow[u, phantom, "\overset{p}{\approx}"] &
\end{tikzcd}$$

Figure 5.4: Diagram for an "approximative" linear relation

The diagram in Figure 5.4 illustrates Formula (3). Note that the linear form κ of the general theory is implicit in the present context: Since the key bits are simply added to plaintext and ("intermediary") ciphertext we have $\kappa = \alpha$ for $k^{(0)}$, and $\kappa = \beta$ for $k^{(1)}$, hence $\kappa(k^{(0)}, k^{(1)}) = \alpha(k^{(0)}) + \beta(k^{(1)})$.

How does this scenario fit the general situation from Chapter 2? In example A we have

- key length $l = 2n$, key space $\mathbb{F}_2^{2n}$, and keys of the form $k = (k^{(0)}, k^{(1)})$ with $k^{(0)}, k^{(1)} \in \mathbb{F}_2^n$.
- The cipher is defined by the map
$$F\colon \mathbb{F}_2^n \times \mathbb{F}_2^n \times \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^n, \quad (a, k^{(0)}, k^{(1)}) \mapsto f(a + k^{(0)}) + k^{(1)}.$$
- The linear form $\kappa\colon \mathbb{F}_2^n \times \mathbb{F}_2^n \longrightarrow \mathbb{F}_2$ is $\kappa(k^{(0)}, k^{(1)}) = \alpha(k^{(0)}) + \beta(k^{(1)})$.

Hence the probability of a linear relation for a fixed key $k = (k^{(0)}, k^{(1)})$ is

$$\begin{aligned}
p_{F,\alpha,\beta,\kappa}(k) &= \frac{1}{2^n} \cdot \#\{a \in \mathbb{F}_2^n \mid \kappa(k) = \alpha(a) + \beta(F(a,k))\} \\
&= \frac{1}{2^n} \cdot \#\{a \in \mathbb{F}_2^n \mid \alpha(k^{(0)}) + \beta(k^{(1)}) = \alpha(a) + \beta(f(a + k^{(0)}) + k^{(1)})\} \\
&= \frac{1}{2^n} \cdot \#\{a \in \mathbb{F}_2^n \mid \alpha(k^{(0)}) = \alpha(a) + \beta(f(a + k^{(0)}))\},
\end{aligned}$$

where we omitted $\beta(k^{(1)})$ that occurs on both sides of the equation inside the curly set brackets.

This expression is independent of $k^{(1)}$, and the slightly rewritten equation

$$p_{F,\alpha,\beta,\kappa}(k) = \frac{1}{2^n} \cdot \#\{a \in \mathbb{F}_2^n \mid \alpha(a + k^{(0)}) = \beta(f(a + k^{(0)}))\}$$

shows that it assumes the same value for all $k^{(0)}$: With a also $a + k^{(0)}$ runs through all of $\mathbb{F}_2^n$ for a fixed $k^{(0)}$. Therefore this value must agree with the mean value over all k:

$$p_{F,\alpha,\beta,\kappa}(k) = p_{F,\alpha,\beta,\kappa} = \frac{1}{2^n} \cdot \#\{x \in \mathbb{F}_2^n \mid \alpha(x) = \beta(f(x))\} = p.$$

This consideration shows:

Proposition 3 *In the scenario of example A the probability $p_{F,\alpha,\beta,\kappa}(k)$ assumes the same value*

$$p = \frac{1}{2^n} \cdot \#\{x \in \mathbb{F}_2^n \mid \alpha(x) = \beta(f(x))\}$$

for all keys $k \in \mathbb{F}_2^{2n}$. In particular p coincides with the mean value from Equation (2).

Using the notations from Figure 5.3 we have

$$\begin{aligned} \beta(c) &= \beta(b' + k^{(1)}) = \beta(b') + \beta(k^{(1)}) \\ &\overset{p}{\approx} \alpha(b) + \beta(k^{(1)}) = \alpha(a + k^{(0)}) + \beta(k^{(1)}) = \alpha(a) + \alpha(k^{(0)}) + \beta(k^{(1)}). \end{aligned}$$

This yields a linear relation for the bits of the key $k = (k_1, k_2)$:

$$\alpha(k^{(0)}) + \beta(k^{(1)}) \overset{p}{\approx} \alpha(a) + \beta(c).$$

Treating the complementary relation

$$\beta \circ f(x) \overset{1-p}{\approx} \alpha(x) + 1$$

in an analoguous way we get:

Proposition 4 *In the scenario of example A let (α, β) be a pair of linear forms for f with probability p as in Formula* (3)*. Then $\hat{p} = \max\{p, 1-p\}$ is the success probability for determing a single key bit by this linear relation given* one *known plaintext block.*

a	b	b'	c	$\alpha(a) + \beta(c)$
0000	1000	0010	0011	1
0001	1001	0110	0111	1
0010	1010	0011	0010	0
0011	1011	0001	0000	1
0100	1100	1001	1000	1
0101	1101	0100	0101	1
0110	1110	0101	0100	1
0111	1111	1000	1001	1
1000	0000	1100	1101	1
1001	0001	1111	1110	1
1010	0010	0111	0110	1
1011	0011	1010	1011	1
1100	0100	1110	1111	1
1101	0101	1101	1100	1
1110	0110	1011	1010	1
1111	0111	0000	0001	0

Table 5.3: A linear relation for the key bits

Example

Take $n = 4$, and for f take the S-box S_0 of LUCIFER. As the two rightmost columns of Table 5.1 show the linear relation defined by (α, β), where $\alpha(x) = x_4$ and $\beta(y) = y_1 + y_2 + y_4$, has probability $p_{f,\alpha,\beta} = \frac{14}{16} = \frac{7}{8}$ (providing strong evidence that the designers of LUCIFER weren't aware of linear cryptanalysis).

As concrete round keys take $k_0 = \mathtt{1000}$ and $k_1 = \mathtt{0001}$. Table 5.3, running through all possible 16 plaintexts, shows that $\alpha(a) + \beta(c)$ assumes the value $1 = \alpha(k_0) + \beta(k_1)$ for this partial sum of key bits exactly 14 times—as expected.

How large is the success probability p_N of correctly estimating this partial sum, assuming $N = 1, 2, \ldots$ random known plaintexts from the set of 2^n possible plaintexts? (For given linear forms α and β with $p = p_{f,\alpha,\beta}$.) This is exactly the scenario of the hypergeometric distribution (for an explanation of the hypergeometric distribution see Appendix E). Therefore we have:

Proposition 5 *In example A let (α, β) be a pair of linear forms that defines a linear relation for f with probability p. Then the success probability for determining a key bit by this linear relation from N known plaintexts is the cumulated probability $p_N = p_N^{(s)}$ of the hypergeometric distribution with parameters 2^n, $s = \hat{p} \cdot 2^n$, and N where $\hat{p} = \max\{p, 1 - p\}$.*

If we neglect exact mathematical reasoning and work with asymptotic

$N\lambda$	1	2	3	4	...	8	9
p_N	84,1%	92,1%	95,8%	97,7%	...	99,8%	99,9%

Table 5.4: Dependence of the success probability on the number of known plaintexts

approximations (as is common in applied statistics), then we can replace the hypergeometric distribution by the normal distribution. The usual (quite vaguely stated) conditions for this approximation are "p not too different from $\frac{1}{2}$, $N \ll 2^n$, but N not too small." This gives the formula

$$p_N \approx \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\sqrt{N\lambda}} e^{-t^2/2}\, dt, \tag{4}$$

where $\lambda = (2p-1)^2$ is the potential of the linear relation. The values associated with the normal distribution are well-known and yield Table 5.4. Instead of the approximation by the normal distribution we could directly use the hypergeometric distribution. This would, in particular for small N, give a more precise value but not a closed formula as simple as (4).

To get a success probability of about 95% we need $N \approx \frac{3}{\lambda}$ known plaintexts according to the table. In the concrete example above we had $p = \frac{7}{8}$, hence $\lambda = \frac{9}{16}$, and the number of known plaintexts needed for a 95% success probability is $N \approx 5$. Using Table 5.2 we succeeded with only $N = 3$ plaintexts. This is not a great surprise because the a-priori probability of this success is about 90% (for $N\lambda = \frac{27}{16} \approx 1,68\ldots$).

> In this example the condition "N not too small" for the approximation by the normal distribution is more than arguable. However determining the exact values for the hypergeometric distribution is easy: Consider an urn containing 16 balls, 14 black ones and 2 white ones, and draw 3 balls by random. Then the probability of all of them being black is $\frac{26}{40}$, the probability of two being black and one being white is $\frac{13}{40}$. Hence the probability of at least two balls being black is $\frac{39}{40} = 97,5\%$. This is clearly more than the 90% from the approximation (4). The remaining probabilities are $\frac{1}{40}$ for exactly one black ball, and 0 for three white balls.