Boolean functions in cryptography

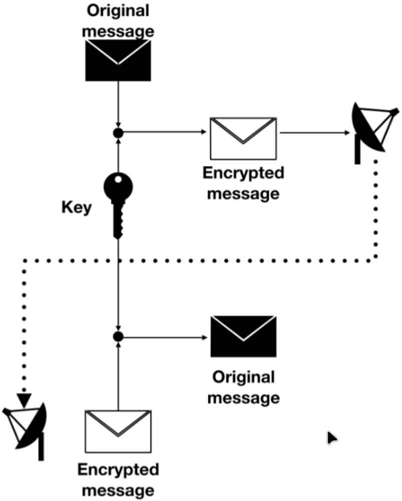
Nikolay Kaleyski

Boolean Seminar Liblice 2023

Boolean functions and vectorial Boolean functions

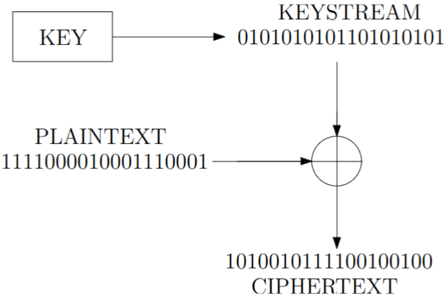
- $\mathbb{F}_2 = \{0, 1\}$;
- $\mathbb{F}_2^n$;
- **Boolean function (BF):** $f : \mathbb{F}_2^n \to \mathbb{F}_2$;
- $\mathbb{F}_{2^n}$;
- $f : \mathbb{F}_{2^n} \to \mathbb{F}_2$;
- **vectorial BF (VBF)**: $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$;
- $F : \mathbb{F}_{2^n} \to \mathbb{F}_{2^m}$;
- also called (n, m)-function;
- $F = (f_1, f_2, \dots, f_m)$ for $f_i : \mathbb{F}_2^n \to \mathbb{F}_2$;
- f_i are the **coordinate functions** of F;
- non-zero linear combinations are the **component functions** of F.

Cryptography and symmetric ciphers



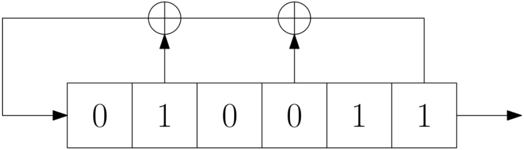
- Secure transmission of a sensitive message across a channel;
- original message = **plaintext**;
- encrypted message = **ciphertext**;
- encryption/decryption only possible with knowledge of the **secret key**;
- symmetric = same key used for encryption and decryption.

Stream ciphers



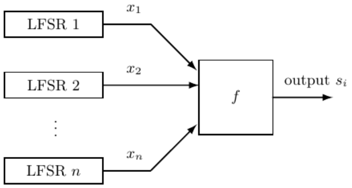
- Plaintext processed as a *stream of bits*;
- short **key** expanded into arbitrarily long **keystream**;
- keystream XOR-ed with plaintext to encrypt;
- keystream XOR-ed with ciphertext to decrypt;
- different stream ciphers = different ways of generating the keystream.

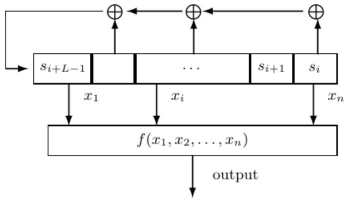
LFSRs



- Linear Feedback Shift Register (LFSR);
- can generate up to $2^n - 1$ states before looping:
  - 01001**1**;
  - 00100**1**;
  - 10010**0**;
  - 11001**0**;
  - 11100**1**;
- insecure due to linear behavior.

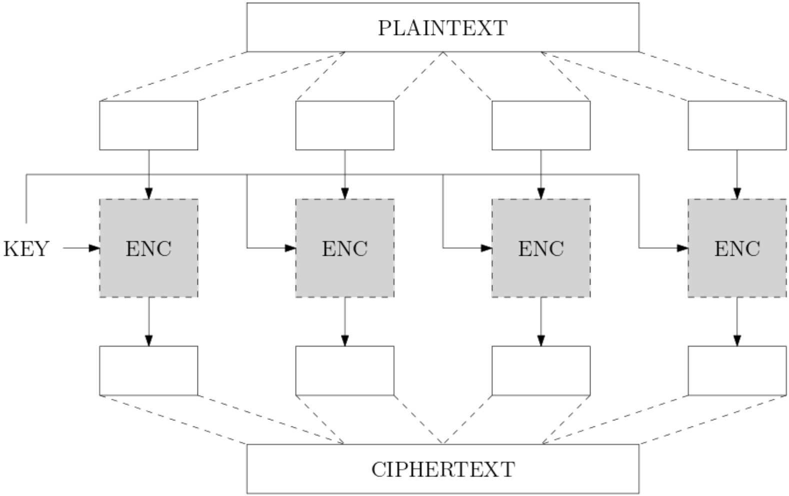
Combiners and filters



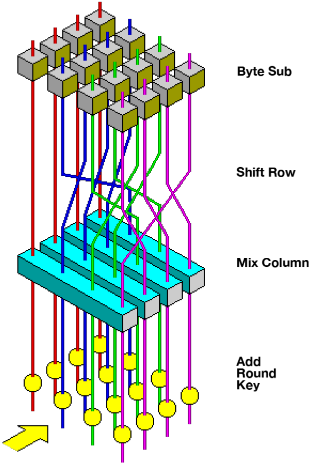


- A **combiner** BF relates the output bits of n LFSRs into a single output bit;
- a **filter** BF computes the output bit as a function of multiple cells;
- the BF is the only nonlinear component of the stream cipher;
- the BF must have "good cryptographic properties".

Block ciphers



Block cipher design



- Data split into smaller blocks;
- interleaved linear operations with a nonlinear VBF;
- round structure repeated multiple times;
- e.g. AES (Advanced Encryption Standard, AKA the Rijndael cipher).

Cryptographic properties

- BFs and VBFs are the only nonlinear parts in ciphers;
- regular structure and patterns can be exploited;
- different attacks = different structure;
- different attacks $\implies$ different properties;
- a good $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ should have near optimal values of all relevant properties;
- finding good tradeoffs is challenging.

Representations

Algebraic Normal Form

x_1	x_2	x_3	$f(x)$
0	0	0	1
0	0	1	0
0	1	0	0
1	0	0	1
0	1	1	1
1	1	0	1
1	0	1	0
1	1	1	1

- $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$;
- $F(x_1, x_2, \ldots, x_n) = \sum_{I \in \mathcal{P}(\{1,2,\ldots,n\})} a_I \prod_{i \in I} x^i$;
- $a_I \in \mathbb{F}_2^m$;
- $f(x_1, x_2, x_3) = 1 + x_2 + x_3 + x_1x_2 + x_1x_2x_3$.

ANF (another example)

x_1	x_2	x_3	x_4	$f_1(x_1, x_2, x_3, x_4)$	$f_2(x_1, x_2, x_3, x_4)$
0	0	0	0	0	0
0	0	0	1	0	1
0	0	1	0	0	1
0	0	1	1	0	0
0	1	0	0	1	0
0	1	0	1	1	1
0	1	1	0	1	1
0	1	1	1	1	0
1	0	0	0	1	0
1	0	0	1	1	1
1	0	1	0	1	1
1	0	1	1	1	0
1	1	0	0	0	0
1	1	0	1	0	1
1	1	1	0	0	1
1	1	1	1	0	0

$$F(x_1, x_2, x_3, x_4) = (1, 0)x_1 + (1, 0)x_2 + (0, 1)x_3 + (0, 1)x_4.$$

Univariate representation

- $\mathbb{F}_2^n \approx \mathbb{F}_{2^n}$;
- $p(x) \in \mathbb{F}_2[x]$ primitive with of degree n;
- $p(\alpha) = 0$;
- $\mathbb{F}_{2^n} = \mathbb{F}_2[\alpha] = \{a_0 + a_1\alpha + \cdots + a_n\alpha^{n-1} : a_0, a_1, \ldots, a_{n-1} \in \mathbb{F}_2\}$;
- $a_0 + a_1\alpha + \cdots + a_n\alpha^{n-1} \approx (a_0, a_1, \ldots, a_{n-1})$;
- $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with $m \mid n$;
- $F : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$;
- $F(x) = \sum_{i=0}^{2^n-1} a_i x^i, a_i \in \mathbb{F}_{2^n}$;
- many important functions have short representations;
- $F(x) = x^3$.

Univariate representation: example

- $F(x) = x^3$ over $\mathbb{F}_{2^3}$;
- $p(x) = x^3 + x + 1$;
- $\alpha^3 + \alpha + 1 = 0$;
- $F(1, 0, 1) =?$;
- $(1, 0, 1) = (1 + \alpha^2)$;
- $(1 + \alpha^2)^3 = (1 + \alpha^4)(1 + \alpha^2)$;
- $= (1 + \alpha + \alpha^2)(1 + \alpha^2)$;
- $= 1 + \alpha^2 + \alpha + \alpha^3 + \alpha^2 + \alpha^4$;
- $= 1 + \alpha + \alpha + 1 + \alpha^2 + \alpha$;
- $F(1, 0, 1) = (0, 1, 1)$.

Balancendess

- BFs should be balanced: $\#f^{-1}(0) = \#f^{-1}(1)$;
- (n, n)-VBFs must be balanced (bijective) for e.g. Substitution-Permutation-Networks;
- it is preferable for VBFs to be balanced in any case;
- $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is balanced if and only if all of its components are balanced.

Algebraic degree

- ANF: $F(x_1, x_2, \dots x_n) = \sum_I a_I \prod_{i \in I} x^i$;
- $\deg(F) = \max\{\#I : a_I \neq 0\}$;
- $\deg(F) \leq 1$: linear (affine) function;
- $\deg(F) = 2$: quadratic;
- $\deg(F)$ should be high;
- low $\deg(F)$ implies a low linear complexity of the output sequence for combiners and filters;
- low $\deg(F)$ allows structural attacks (integral, cube, higher-order differential) in block ciphers.

Nonlinearity

- How well $f : \mathbb{F}_2^n \to \mathbb{F}_2$ can be approximated by an affine function $a \in \mathcal{A}_n = \{a : \mathbb{F}_2^n \to \mathbb{F}_2, \deg(F) \leq 1\}$;
- $\mathcal{NL}(f) = \min\{d_H(f, a) : a \in \mathcal{A}_n\}$;
- low $\mathcal{NL}(f)$ allows fast correlation attacks for stream ciphers;
- covering radius bound: $\mathcal{NL}(f) \leq 2^{n-1} - 2^{n/2-1}$;
- $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is **bent** if $\mathcal{NL}(f) = 2^{n-1} - 2^{n/2-1}$.

Other properties

- $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is t**-th order correlation immune (CI)** if its output distribution is unaltered when at most t bits are fixed;
- if $fg = 0$, g is called an **annihilator** of f;
- **algebraic immunity** $AI(f)$ = lowerst algebraic degree of an annihilator of f or $(f + 1)$;
- $D_a f(x) = f(a + x) + f(x)$;
- f satisfies the **propagation criterion (PC)** w.r.t. E if $D_a f$ is balanced for all $a \in E$;
- if $D_a F$ is constant, then a is a **linear structure** of f;
- linear structures should not exist;
- ...

Differential uniformity

- $D_a F(x) = F(a + x) + F(x)$;
- $D_a F(x) = b$;
- $\delta_F(a, b) = \#\{x \in \mathbb{F}_2^n \mid D_a F(x) = b\}$;
- $\Delta_F = \max\{\delta_F(a, b) : 0 \neq a, b\}$;
- low Δ_F gives good resistance to differential attacks;
- always even;
- if $\Delta_F = 2$, then F is almost perfect nonlinear (APN).

Research directions

- Infinite families (constructions of good e.g. (n, n)-functions for infinitely many n);
- e.g. $F(x) = x^3 + \beta x^{2^i-1} + \beta^2 (x^3)^{2^{n/2}} + (x^{2^i-1})^{2^{n/2}}$ is APN over $\mathbb{F}_{2^n}$ for $n = 10 + 4k$ for $i = n/2 - 1$ or $i = (n/2-1)^{-1} \pmod{n}$;
- computational searches for good functions;
- for example, for functions with a simple form under some representation;
- efficiently testing properties.

Equivalence relations

- Classes of cryptographic functions considered only up to "appropriate" equivalence relations;
- CCZ-equivalence (Carlet-Charpin-Zinoviev): $L(\Gamma_F) = \Gamma_G$ where $\Gamma_F = \{(x, F(x)) : x\}$ and L is a linear permutation;
- EA-equivalence (extended affine): $A_1 \circ F \circ A_2 + A = G$, where A_1, A_2, A affine and A_1, A_2 permutations;
- affine equivalence: $A = 0$;
- linear equivalence: $A_1(0) = A_2(0) = 0$;
- testing equivalence is also hard!

Thank you for your attention!