Highly Nonlinear Boolean Functions with Optimal Algebraic Immunity

Claude Carlet

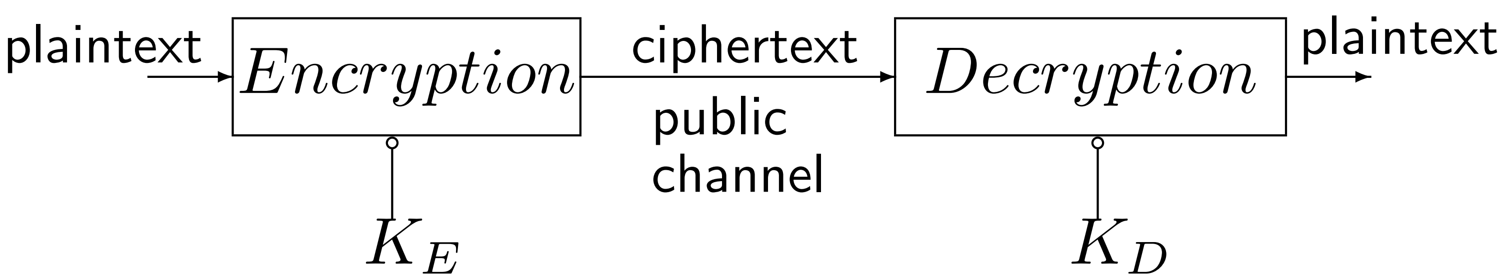
LAGA, Universities of Paris 8 and Paris 13, CNRS, France

Outline

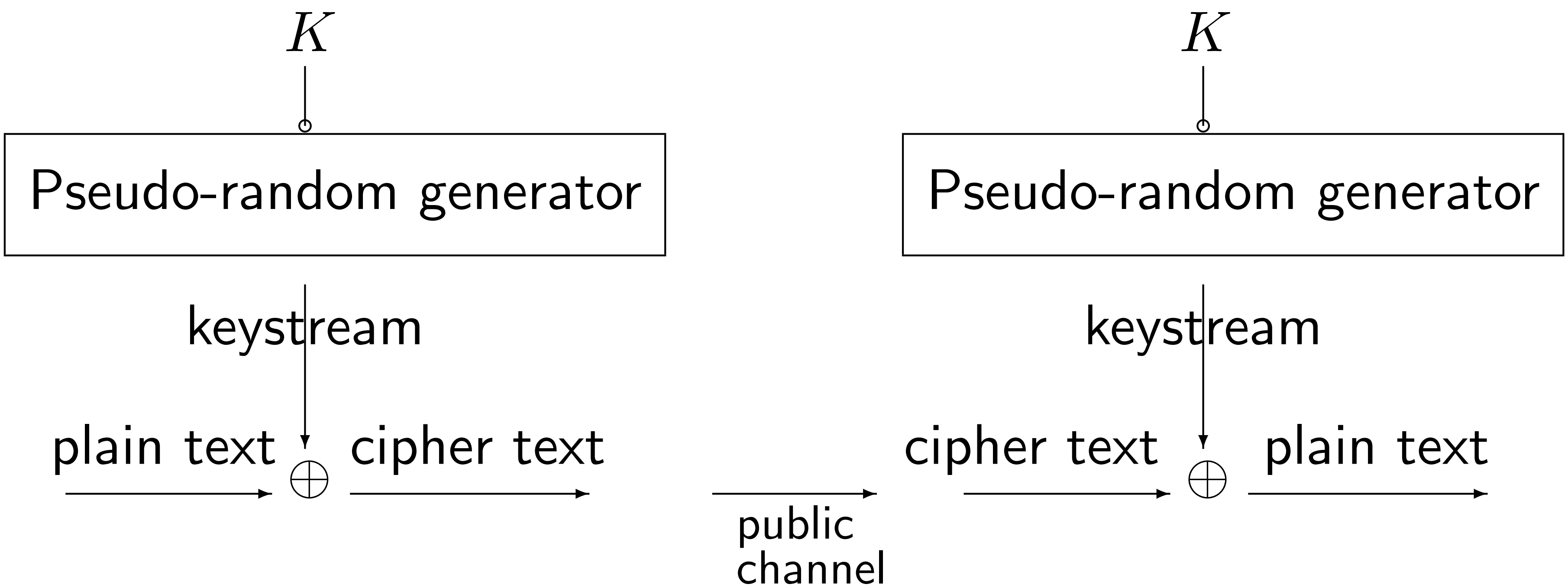
- Preliminaries on stream ciphers and Boolean functions
- Algebraic attacks on stream ciphers and algebraic immunity
- The known Boolean functions with optimal algebraic immununity
- Recent developments

Preliminaries on stream ciphers and Boolean functions

Ciphers (cryptography) :



Synchronous stream ciphers :



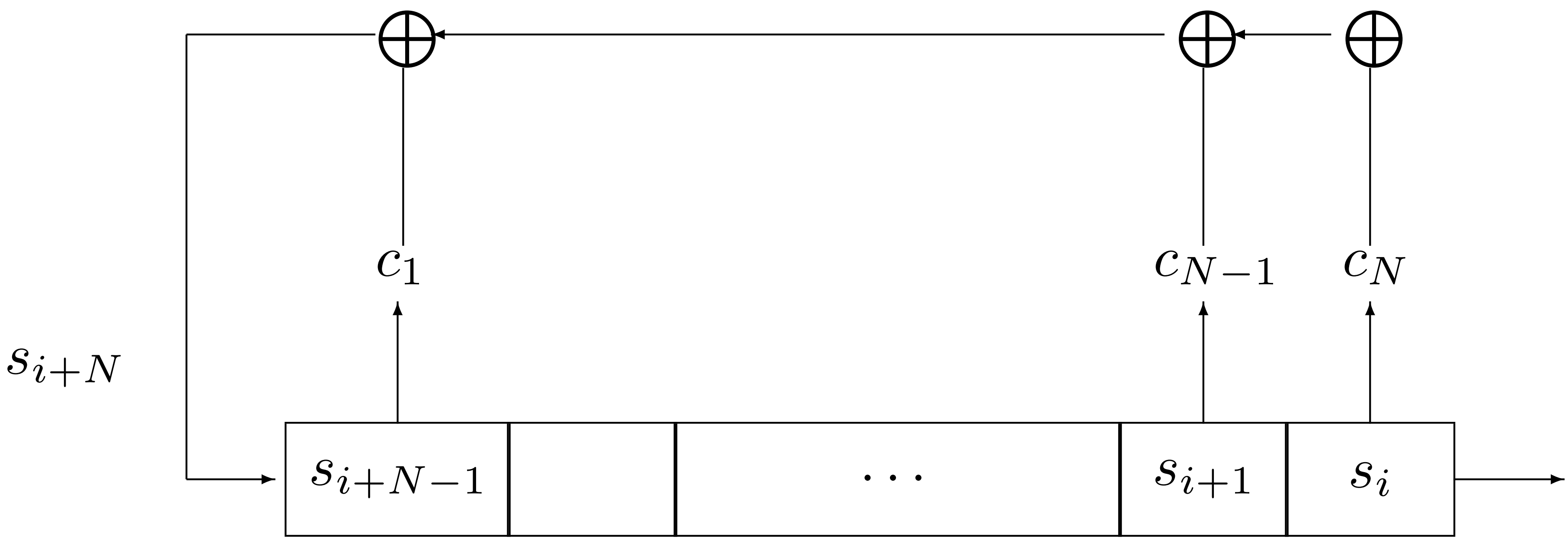
Every PRG consists in a linear part (for efficiency) and a nonlinear part (for robustness).

Boolean functions $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ are often used in the nonlinear part.

There exist **two theoretical models** for their use in the pseudo-random generators (PRG) of Synchronous stream ciphers.

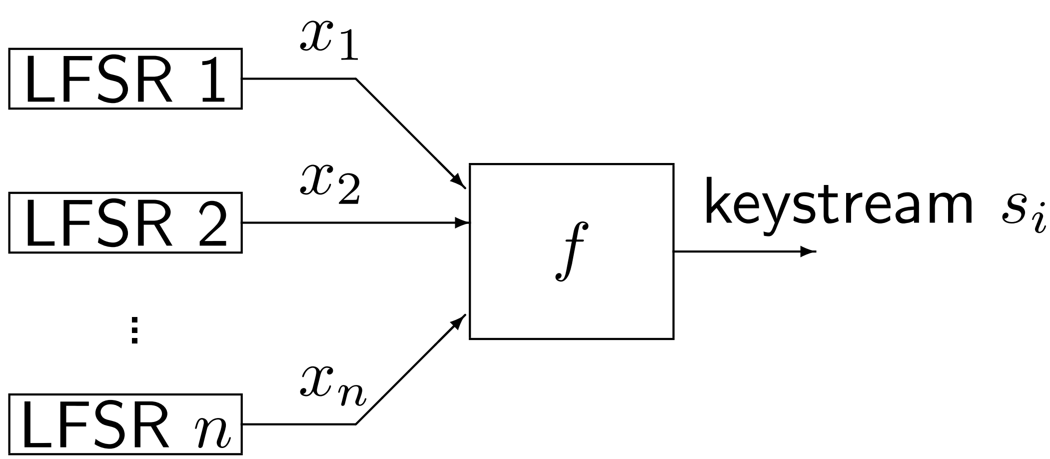
Both use Linear Feedback Shift Registers in the linear part :

Linear feedback shift registers :

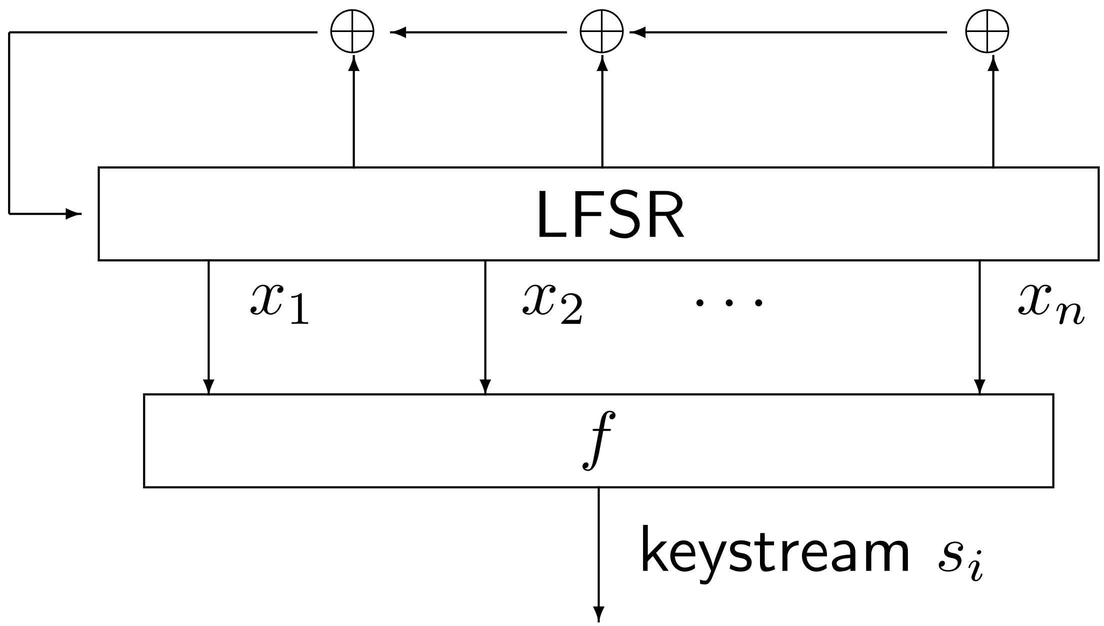


$s_i = \sum_{j=1}^{N} c_j s_{i-j}$.

Combiner model :



Filter model



In both models, f must be balanced to avoid distinguishing attacks.

Two representations of Boolean functions :

- *The Algebraic Normal Form* (ANF) :

$$f(x_1, \cdots, x_n) = \sum_{I \subseteq \{1,\ldots,n\}} a_I \left(\prod_{i \in I} x_i \right), \ a_I \in \mathbb{F}_2.$$

The ANF exists and is unique.

The algebraic degree is the degree of the ANF.
It must be large because of Berlekamp-Massey and Rønjom-Helleseth attacks.

Affine functions : sums of linear functions and constants :
$a_1\, x_1 + \cdots + a_n\, x_n + \epsilon = a \cdot x + \epsilon$; $\quad a \in \mathbb{F}_2^n; \quad \deg \leq 1.$
Their set is the Reed-Muller code of order 1.

- *The univariate representation (the trace representation)* :

- The vector space $\mathbb{F}_2^n$ is endowed with the structure of the field $\mathbb{F}_{2^n}$.
Any function $f : \mathbb{F}_{2^n} \mapsto \mathbb{F}_{2^n}$ admits the unique representation :

$$f(x) = \sum_{j=0}^{2^n-1} a_j\, x^j; \quad a_j,\, x \in \mathbb{F}_{2^n}.$$

- f is Boolean if and only if :

$$a_0, a_{2^n-1} \in \mathbb{F}_2 \text{ and } a_{2j} = (a_j)^2, \forall j \in \mathbb{Z}/(2^n-1)\mathbb{Z}.$$

Hence :

$f(x) = tr(P(x))$, where $tr(x) = x + x^2 + x^{2^2} + \cdots + x^{2^{n-1}}$.

Then the algebraic degree equals : $\max\{w_2(j);\ j$ s.t. $a_j \neq 0\}$, where $w_2(j)$ is the Hamming weight of the binary expansion of j.

Affine functions $tr(ax) + \epsilon$, $a \in \mathbb{F}_2^n$, $\epsilon \in \mathbb{F}_2$.

The *Walsh transform* of a Boolean function :

$$\widehat{\mathrm{f}}(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+a\cdot x} \text{ or } \sum_{x \in \mathbb{F}_{2^n}} (-1)^{f(x)+tr(ax)}.$$

The *Hamming distance* between two functions :

$$d_H(f,g) = w_H(f+g) = |\{x \in \mathbb{F}_2^n \,/\, f(x) \neq g(x)\}|.$$

The *nonlinearity* of a Boolean function f is the minimum Hamming distance from f to affine functions (i.e. its distance to the Reed-Muller code of order 1) and equals :

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} |\widehat{\mathrm{f}}(a)|.$$

The nonlinearity nl is upper bounded by $2^{n-1} - 2^{n/2-1}$ (covering radius bound). This maximum is achieved by bent functions.

The nonlinearity nl must be large to prevent the system from fast correlation attacks.

Balancedness, high algebraic degree and large nonlinearity was considered as roughly sufficient for the filter model of pseudo-random generator before the introduction of algebraic attacks.

Algebraic attacks on stream ciphers and algebraic immunity

Algebraic attacks : *Principle* (Shannon) :

-Find equations with the key bits as unknowns
-Solve the system of these equations.

For stream ciphers (combiner model and filter model) :
- denote by $(s_0, \ldots, s_{N-1})$ the initial state of the linear part of the pseudo-random generator ;
- there exists a linear automorphism L and a linear mapping L' s.t.

$$s_i = f(L' \circ L^i(s_0, \ldots, s_{N-1})).$$

Problem of the general algebraic attack :

Highly non-linear equations with many unknowns.

But with stream ciphers we can have many equations $\rightarrow$

over-defined system.

One can then linearize the system (or use Gröbner bases).

However the number of unknowns is then much too large.

Courtois-Meier : If one can find $g \neq 0$ and h of low degrees such that $fg = h$, then the equation $s_i = f(L' \circ L^i(s_0, \ldots, s_{N-1}))$ implies the low degree equation :

$$s_i \, g(L' \circ L^i(s_0, \ldots, s_{N-1})) = h(L' \circ L^i(s_0, \ldots, s_{N-1}))$$

and the degree of the nonlinear system and the number of unknowns in the related linear system decrease.

Algebraic immunity :

A necessary and sufficient condition for existence of low degree $g \neq 0$ and h such that $fg = h$ (Meier-Pasalic-C.C.) :
there exists $g \neq 0$ of low degree such that $fg = 0$ or $(f + 1)g = 0$.

Definition : a function g such that $fg = 0$ is called an $annihilator$. The $algebraic\ immunity$ $AI(f)$ is the minimum degree of the nonzero annihilators of f and of those of $f + 1$.

Related to coding problems over the erasure channel.

We have : $AI(f) \leq \deg(f)$ and $AI(f) \leq \lceil \frac{n}{2} \rceil$.

A variant of algebraic attacks, called "fast algebraic attack" needs the existence of $g \neq 0$ and h such that $fg = h$, where only g has low degree and h has reasonable degree.

The known Boolean functions with optimal algebraic immununity

2000-2005 :

- The majority function defined by :

$$f(x) = 1 \text{ iff } w_H(x) \geq n/2.$$

and its generalizations by Dalai et al., Bracken, C.C... ;

- An iterative construction (Dalai-Gupta-Maitra), n even.

These functions have high degree but *insufficient nonlinearity* and bad resistance to Fast Algebraic Attacks (Dalai, Gupta, Maitra, Armknecht, C.C., Gaborit, Meier, Ruatta...).

2008 :

Definition [CF function]

Let $n \geq 2$ *and* α *a primitive element of the field* $\mathbb{F}_{2^n}$*.*
We denote by f *the Boolean function on* $\mathbb{F}_{2^n}$ *whose support is* $\{\alpha^s, \cdots, \alpha^{2^{n-1}+s-1}\}$*.*

Theorem (Feng, Liao, Yang)

The function f *defined above has optimal algebraic immunity* $\lceil n/2 \rceil$*.*

Better proof (sketch) by C.C., Feng :

Let $g(x) = \sum_{j=0}^{2^n-1} a_j\, x^j$ be a non-zero annihilator of $f+1$.

g is a codeword of a Reed-Solomon code of designed distance $2^{n-1}+1$.

Hence $|\{j \,/\, a_j \neq 0\}| \geq 2^{n-1}+1$ and $\deg(g) \geq \lceil \frac{n}{2} \rceil$.

Algebraic degree (C.C., Feng) : f has degree $n-1$ (optimal).

Nonlinearity (C.C., Feng) :

$$nl(f) \geqslant 2^{n-1} - \frac{2^{\frac{n}{2}+1}}{\pi} \ln\left(\frac{4(2^n-1)}{\pi}\right) - 1 \;\sim\; 2^{n-1} - \frac{\ln 2}{\pi}\, n\, 2^{\frac{n}{2}+1}.$$

The function behaves well against fast algebraic attacks for small values of n.

Recent developments

Definition (Z. Tu and Y. Deng - Designs, Codes and Cryptography)

$$(x,y) \in \mathbb{F}_{2^n} \times \mathbb{F}_{2^n};\ f^{\#}(x,y) = f(xy^{2^n-2}) = f\left(\frac{x}{y}\right),\ \text{with } \frac{x}{0} = 0.$$

Theorem (Z. Tu and Y. Deng) <u>*up to a conjecture*</u>

The function $f^{\#}$ has optimal algebraic immunity n.

Nonlinearity :

$$nl(f^{\#}) = 2^{2n-1} - 2^{n-1}$$

($f^{\#}$ has best possible nonlinearity ; it is bent).

Remark. Function $f^{\#}$ is not balanced and has degree at most n (as any bent function). But the function :

$$f^{\#'}(x,y) = \begin{cases} f\left(\frac{x}{y}\right) \text{ if } y \neq 0 \\ f(x) \text{ if } y = 0 \end{cases}$$

has optimal algebraic immunity as well and is balanced. Its degree equals $2n-1$ and $nl(f^{\#'}) \geq 2^{2n-1} - 2^{n-1} - n\, 2^{n/2} \ln 2 - 1$.

But observations :

- This function is weak against the fast algebraic attack (C.C., IACR ePrint Archive).

- Its distance to functions of algebraic degrees at most $n/2$ is small and this implies that its resistance to fast algebraic attack is weak (Wang-Johansson, INSCRYPT 2010).

Any function constructed with a similar method would have the same drawback.

Definition (D. Tang, C.C., X. Tang)

$$n \geq 2;\ (x, y) \in \mathbb{F}_{2^n} \times \mathbb{F}_{2^n}; \qquad f_1(x, y) = f(xy).$$

Algebraic immunity : $AI(f_1) = n$.

Immunity to fast algebraic attacks :

Computer exhaustive investigation of g_1, h_1 such that $1 \leq \deg(g_1) = e < n$, $\deg(h_1) = d$ and $f_1 * g_1 = h_1$:

- For $n = 4, 6, 7, 8$, we only found pairs (e, d) such that $e + d \geq 2n - 2$, when $1 \leq e < n$.
- For $n = 2, 3, 5$, we only found pairs (e, d) such that $e + d \geq 2n$, when $1 \leq e < n$.

Recall that for any n-variable function f there exist a pair of functions $(g_1, h_1) \in \mathcal{B}_n \times \mathcal{B}_n$ of respective degrees $1 \leq e < \lceil n/2 \rceil$ and $d = n - e$ such that $f * g_1 = h_1$.

Therefore f_1 has optimal immunity to fast algebraic attacks for $n = 4, 6, 10$ and nearly optimal immunity for $n = 8, 12, 14, 16$.

Algebraic degree : $2n - 2$.

Nonlinearity : $N_{f_1} > 2^{2n-1} - \big(\frac{\ln 2}{\pi} n + 0.42\big) 2^n - 1$.

Slight modification to get balanced functions :

$$f_2(x, y) = \begin{cases} f_1(x, y), & x \neq 0 \\ u(y), & x = 0 \end{cases}$$

where u is balanced on $\mathbb{F}_{2^n}$ satisfying $u(0) = 0$, $\deg(u) = n - 1$, and $\max_{a \in \mathbb{F}_{2^n}} |W_u(a)| \leq 2^{\frac{m+1}{2}}$ if $t = 1$ and $\max_{a \in \mathbb{F}_{2^k}} |W_u(a)| \leq \sum_{i=1}^{t-1} 2^{\frac{n}{2^{i+1}}} + 2^{\frac{m+1}{2}}$ if $t \geq 2$ (u does exist).

Algebraic degree and algebraic immunity f_2 has maximal algebraic degree for balanced function and optimal algebraic immunity.

Immunity to fast algebraic attacks, Nonlinearity : similar to f_1

The exact values of the nonlinearity

n	4	6	8	10	12	14
$2^{n-1}-2^{n/2}$	4	24	112	480	1984	8064
$\mathcal{N}_{CF}$	4	24	112	484	1970	8036
$\mathcal{N}_{f_2}$	4	22	108	476	1982	8028

n	16	18	20	22	24	26
$2^{n-1}-2^{n/2}$	32512	130560	523264	2095104	8384512	33546240
$\mathcal{N}_{CF}$	32530	130442	523154	2094972	8384536	33545716
$\mathcal{N}_{f_2}$	32508	130504	523144	2094980	8384490	33545992

n	28	30	32	34	36	38
$2^{n-1}-2^{n/2}$	134201344	536838144	2147418112	8589803520	34359476224	137438429184
$\mathcal{N}_{f_2}$	134201460	536838052	2147416552	8589818968	34359469052	137438441620