Highly nonlinear filter Boolean functions with high algebraic immunity for stream ciphers

Claude Carlet

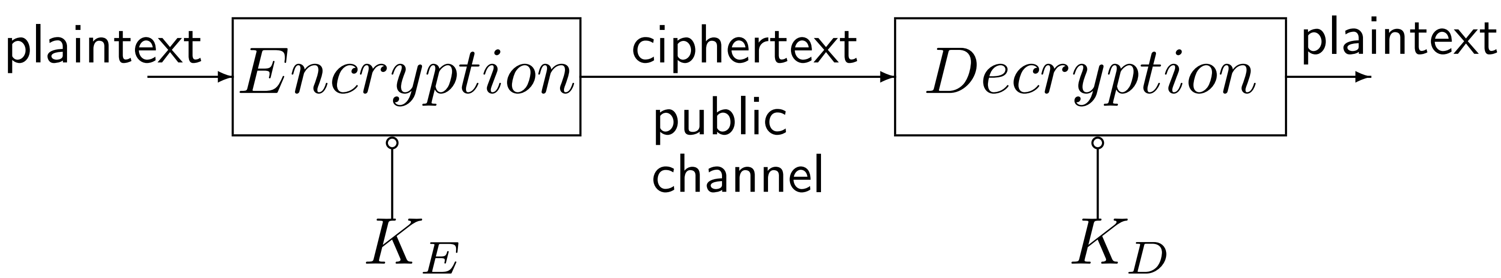
University of Paris 8, France

Outline

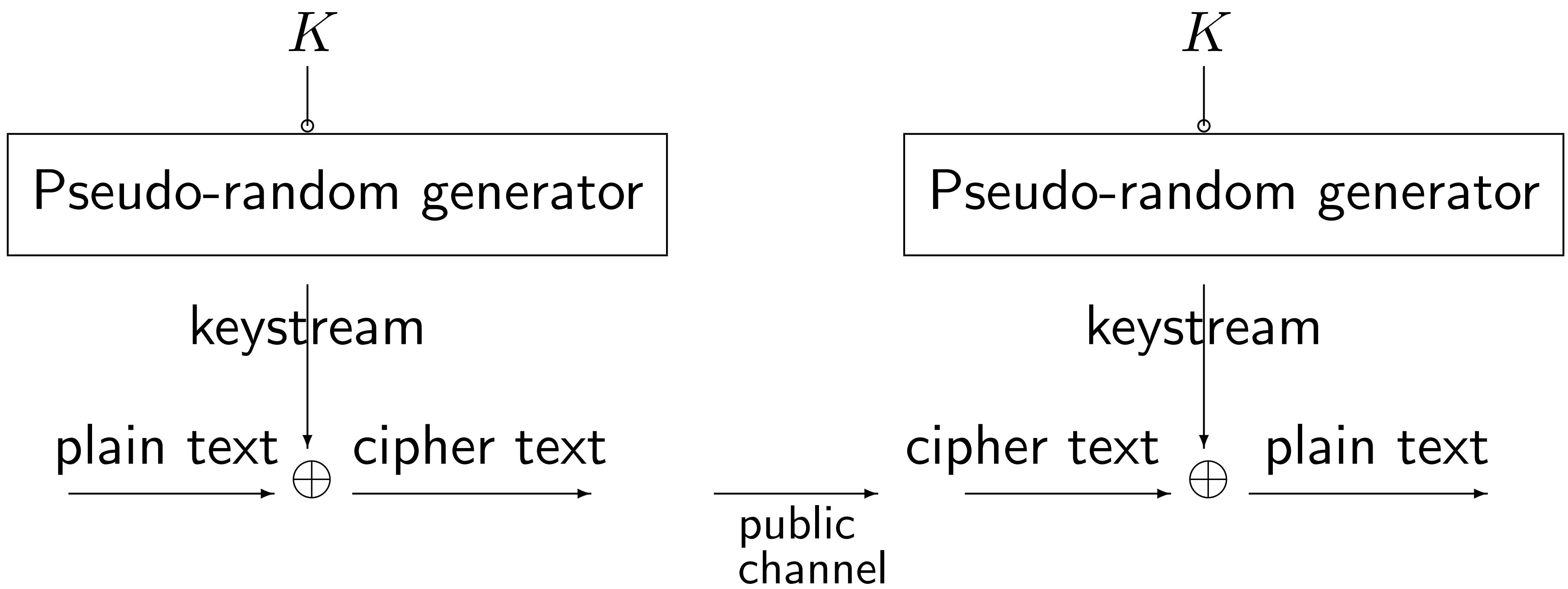
- Preliminaries on stream ciphers and Boolean functions
- Algebraic attacks on stream ciphers and algebraic immunity
- The known Boolean functions with optimal algebraic immununity
- Recent developments

Preliminaries on stream ciphers and Boolean functions

Ciphers (cryptography) :



Synchronous stream ciphers :



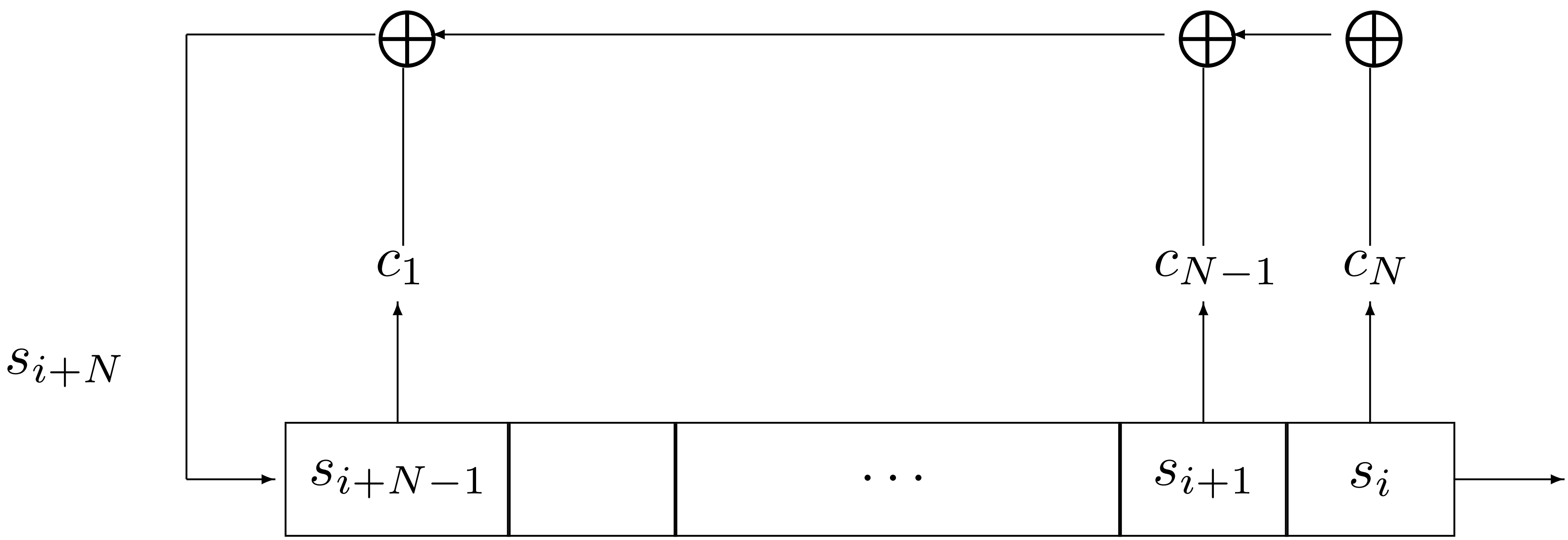
Every PRG consists in a linear part (for efficiency) and a nonlinear part (for robustness).

Boolean functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ 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.

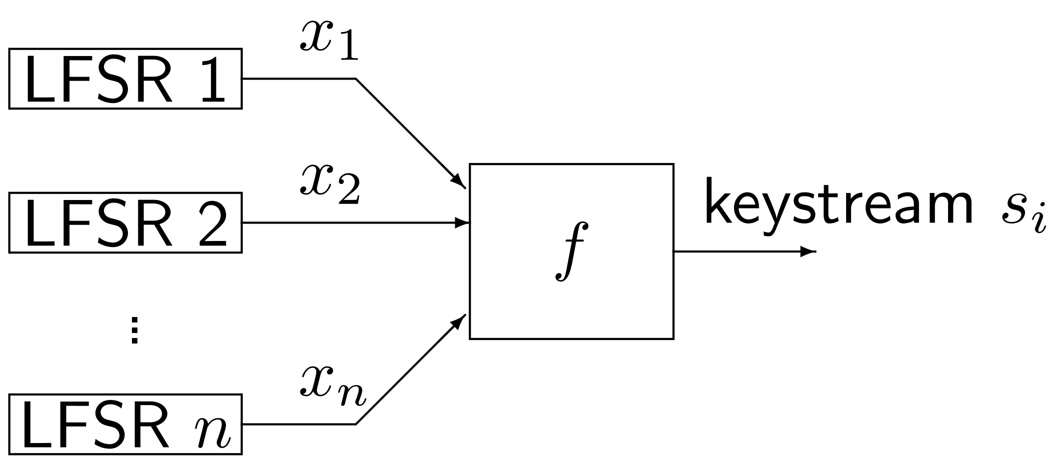
Both use Linear Feedback Shift Registers in the linear part :

Linear feedback shift registers :

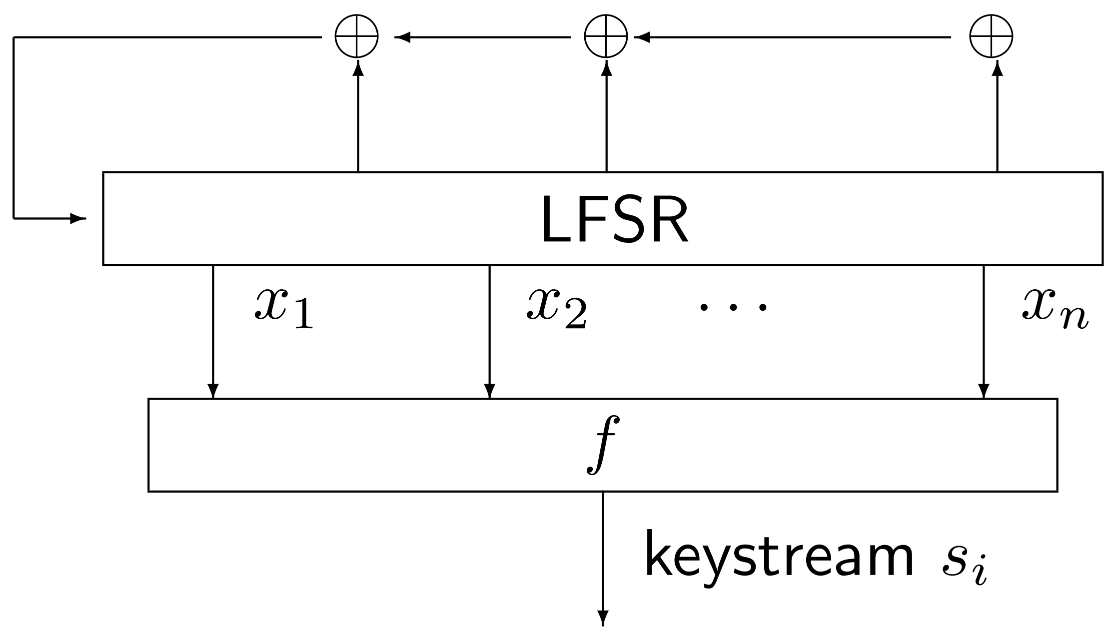


$s_i = \sum_{j=1}^{N} c_j s_{i-j}$.

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 :
$a_1 x_1 + \cdots + a_n x_n + \epsilon = a \cdot x + \epsilon$; $a \in \mathbb{F}_2^n$; $\deg \leq 1$.
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 $AI(f) \leq \lceil \frac{n}{2} \rceil$.

In practical situation, $AI(f)$ *must be greater than or equal to 7*.

Hence we need $n \geq 13$ and in fact $n \approx 20$.

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

Until recently, two classes existed :

- 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.

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

A recently found infinite class of balanced functions with optimal algebraic immunity :

Definition

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* $\{0, 1, \alpha, \cdots, \alpha^{2^{n-1}-2}\}$*.*

Theorem (Feng, Liao, Yang - C.C., Feng)

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

Proof (sketch) :

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) :

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

Exact values of the nonlinearity for f :

n	6	7	8	9	10	11
Best nl known before	22	48	98	196	400	798
The exact values of nl	*24*	*54*	*112*	*232*	*478*	*980*
upper bound $2^{n-1} - 2^{n/2-1}$	28	58	120	244	496	1001

The function seems to behave well against fast algebraic attacks.

The problem of computing the output to the function :

The complexity of computing $f(x)$ is same as for the discrete log. But n is “small”.

The complexity is lower when n is even.

We can then use the Pohlig-Hellman method, with tables for the discrete log for the sizes $2^{n/2} - 1$ and $2^{n/2} + 1$.

The time for computation will be very reasonable (1 bit per cycle) but this will need about 200,000 transistors for $n = 20$.

It is possible to reduce the number of transistors by cutting in three pieces instead of two :

$$2^{18} - 1 = 27 * 73 * 133 \; ; \quad 2^{20} - 1 = 41 * 93 * 275.$$

This reduces the number of transistors to 40,000 for $n = 20$.

Recent developments

Definition (Z. Tu and Y. Deng - IACR ePrint archive)

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

Theorem (Z. Tu and Y. Deng) *up to a conjecture*

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

Related conjecture (checked by Z. Tu and Y. Deng til $n = 29$) :

$\forall n \geq 1, \forall t \in \mathbb{Z}/(2^n - 1)\mathbb{Z}$:

$$card\,\{i \in \mathbb{Z}/(2^n - 1)\mathbb{Z} \,|\, w_2(i) + w_2(t - i) \leq n - 1\} \leq 2^{n-1}.$$

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 observation (C.C.) :

This function is weak against the fast algebraic attack.

It seems difficult to modify it so that it resists FAA.

Conclusion

There exists only one infinite class of functions which potentially satisfies all the necessary criteria for being used as a filter function.

But proving its good behavior is a twofold open problem.

Finding such proof or discovering new classes provably satisfying all the necessary criteria is vital for the future of the filter model.

Announcement : Next SETA conference

Sequences and Their Applications

will be held in ***Paris*** in ***September 13-17, 2010***.

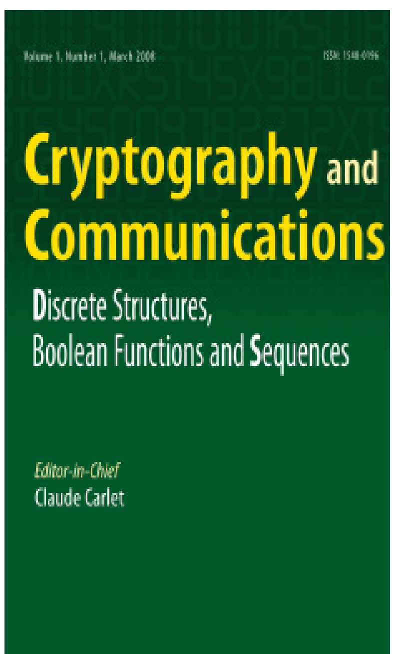
General chair : Patrick Solé

PC co-chairs : A. Pott ; C. Carlet

Springer
the language of science

springer.com

Call for Papers



Cryptography and Communications

Discrete Structures, Boolean Functions and Sequences

A new international journal bridging the domains of coding, cryptography, and communications

28

Editor-in-Chief
Claude Carlet, *University of Paris 8, France*

Cryptography and Communications

Discrete Structures, Boolean Functions and Sequences

First papers by :

Harald Niederreiter, Charles J. Colbourn, Joan Daemen, Vincent Rijmen, Andrew Klapper, Thomas Johansson, Wilfried Meidl...

Numbers of submissions :
- 2008 : 12
- 2009 : 51 (including a special issue)
- 2010 : 30 (up to now).

Success rate : 30%

Book : Cambridge University Press

Boolean Models and Methods in Mathematics, Computer Science, and Engineering

Series : Encyclopedia of Mathematics and its Applications

Edited by *Yves Crama and Peter L. Hammer*

Available in august 2010

8. Boolean functions for cryptography and error-correcting codes, pages 257-397

9. Vectorial Boolean functions for cryptography, pages 398-469