Diffusion in Cryptography

Laurent Poinsot and Sami Harari

Université de Toulon et du Var, I.S.I.T.V.
Laboratoire S.I.S.
BP 132
83 957 La Garde cédex, France
{laurent.poinsot,sami.harari}@univ-tln.fr

Abstract. The diffusion of information into functions involved in secret-key cryptography is a crucial criterion for robustness of algorithms against some statistical attacks such as differential and linear cryptanalysis. Many authors have used their own definition for diffusion in order to formalize the properties of solidity of their cryptosystems. These notions, although more or less distinct, share the same objective : the quantitative study of the correlation between the amount of information in input and in output of a function. This paper is a summary of several of these notions and an attempt to explicit the underlying concepts of the general term diffusion.

1 Introduction

During a long time, the robustness of algorithms involved in secret-key cryptography relied only on the use of large cardinality sets of boolean functions. Recently however, some works have highlighted new criteria of solidity : in particular, the non-linearity of functions, in order to avoid linear cryptanalysis, the resistance against differential cryptanalysis and the high diffusion of bits of information along functions.
Computable functions of several variables are, in mostly cases, a mathematical product of functions, each of them depending on few variables. This property of "local calculation" leads to *divide and conquer* algorithms to compute such functions. The high diffusion functions are *a contrario* functions which should not exhibit such decompositions. More generally, the diffusion of information denotes the way that effects on inputs of a given function are reflected on its outputs.

Our present work is thus motivated by the study of high diffusion functions in order to improve the robustness of cryptographic algorithms and more generally the study of the diffusion of information. This paper takes the form of a synthesis of several approaches of the problem of diffusion by different authors. This synthesis is in no way exhaustive but presents some interesting notions upon the diffusion.

This paper is organized as follows : first, we present the concept of diffusion such as Shannon has initially introduced it. Then we establish and gather several definitions as a basis for diffusion in the context of boolean functions. In the third part, we present a work from Massey using graph theory. Our synthesis is ended by the description of diffusion given in the design of Rijndael.

2 Historical introduction of Diffusion

The basics of diffusion of information were introduced by C. Shannon in one of his famous papers [1] published in 1949, on secrecy systems. He described the method of diffusion to design cryptosystems in order to resist statistical attacks, according to the following terms : "*... the statistical structure* (of the set of plaintexts) *which leads to its redundancy is "dissipated" into long range statistics - i.e., into statistical structure involving long combination of letters into the cryptogram*". The goal of this method is to scatter the most frequent patterns of plaintexts into long parts of ciphertexts or also to distribute the statistical influence of individual letters over several letters after ciphering. Then the discovery of interesting statistics on plaintexts with knowledge on ciphertexts becomes more difficult. In order to achieve the diffusion, Shannon used enciphering functions considered as channels *with finite input memory and finite anticipation* between the sources of plaintexts and ciphertexts.
In this part, we present the *diffusion channels* of information in the probabilistic way given by Shannon. Let $(\Omega, \mathscr{F}, P)$ be a probability space and $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ two measurable spaces. We denote by X and Y two random variables from $(\Omega, \mathscr{F}, P)$ to respectively $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$. We denote by P_X the measure of probability induced by X on $(\Omega_1, \mathscr{F}_1)$, $P_{X|Y}$ the (family of) conditional measure(s) on $(\Omega_1, \mathscr{F}_1)$ of X given Y and by P_{XY} the joint probability measure on the product space $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2)$. Let $\mathcal{T}$ be a set (the *time*). We denote by $(\times_{t\in\mathcal{T}}\Omega_t, \times_{t\in\mathcal{T}}\mathscr{G}_t)$ the generalized product space of sequences of time $\mathcal{T}$ where for all $t \in \mathcal{T}$, $(\Omega_t, \mathscr{G}_t)$ is a measurable space. In the case where all $(\Omega_t, \mathscr{G}_t)$ are equal to $(\Omega, \mathscr{G})$, we denote the product space simply by $(\Omega^{\mathcal{T}}, \mathscr{G}^{\mathcal{T}})$. A *discrete source of information with alphabet* A is a random variable $X = \{X_n\}_{n\in\mathcal{T}}$ from $(\Omega, \mathscr{F}, P)$ to $(A^{\mathcal{T}}, \mathscr{G}^{\mathcal{T}})$ where $\mathcal{T}$ is a discrete set and A is a finite set.
A natural language can be identified with a discrete source of information (with alphabet the set of letters of the language) which produces each time a letter of a word of the language according to given probabilities. Thus Shannon represented the set of plaintexts as a discrete source of information $\{X_n\}_{n\in\mathbb{N}}$. Moreover the enciphering function applied on messages can be represented by a *discrete channel of communication* which produces one letter of ciphertexts (the sets of ciphertexts is then also a source of information) by respect to several letters of plaintexts. Formally a discrete channel of communication with input alphabet A and output alphabet B is a 5-uplet $((\Omega, \mathscr{F}, P), (A, \mathscr{G}_A), (B, \mathscr{G}_B), X, P_{Y|X})$ where $X = \{X_n\}_{n\in\mathbb{N}}$ is the *input* (discrete) source of information with alphabet A (*i.e.*, X is a random process from $(\Omega, \mathscr{F})$ to $(A^{\mathbb{N}}, \mathscr{G}_A^{\mathbb{N}})$) and Y is the *output* (discrete) source of alphabet B and determined by the conditional probability

$P_{Y|X}$.
According to Shannon the diffusion is provided by *diffusion channels* which are defined as discrete channels with finite input memory and finite anticipation. Let denote the finite sets of the form $\{n,\ n+1,\ \ldots,\ n+m\}$ by $\{\{n,\ n+m\}\}$ and the infinite sets $\{n, n+1, \ldots\}$ by $\{\{n,\ +\infty\}\}$. By *finite input memory*, we refer to discrete channels of communication such that there exists $M \geq 1$ and for all $n \geq M$ and for all $F \in \mathscr{G}_B^{\{\{n,\ +\infty\}\}}$, $P_{Y|X}(B^{\{\{0,\ n-1\}\}} \times F|\{x\}) = P_{Y|X}(B^{\{\{0,\ n-1\}\}} \times F|\{x'\})$ for all $(x, x') \in (A^{\mathbb{N}})^2$ such that $x_i = x'_i\ \forall i \geq n - M$. We note that $P_{Y|X}(B^{\{\{0,\ n-1\}\}} \times F|\{x\}) = P(\{\omega \in \Omega|(Y_n(\omega), Y_{n+1}(\omega),\ \ldots) \in F\}|X^{-1}(\{x\}))$. In other terms, for an event involving Y_i after some time n, the only past inputs which determine the output probability are the ones for the same time and M time units earlier.
With finite anticipation, we define the following property of discrete channels : there exists an integer L such that for all n and all $F \in \mathscr{G}_B^{\{\{0,n\}\}}$, $P_{Y|X}(F \times B^{\{\{n+1,+\infty\}\}}|\{x\}) = P_{Y|X}(F \times B^{\{\{n+1,+\infty\}\}}|\{x'\})$ for all $(x, x') \in (A^{\mathbb{N}})^2$ such that $x_i = x'_i\ \forall i \leq n+L$. Since $P_{Y|X}(F \times B^{\{\{n+1,+\infty\}\}}|\{x\}) = P(\{\omega|(Y_0(\omega),\ \ldots,\ Y_n(\omega)) \in F\}|X^{-1}(\{x\}))$, only L future inputs must be known to determine the probability of an event involving current and past outputs.

The notion of diffusion given by Shannon is based on expected properties of channels but not on a constructive way even if it has been explicitly used in the design of DES with permutation boxes. As an applied science, we are interested in less abstract criterion which could be deduced from given cipher functions. However we have to keep in mind the historical definition of diffusion. In the next section, we give a boolean function approach of the problem of diffusion.

3 Diffusion of Information of Boolean Functions

The notion of Shannon for the diffusion has been interpreted by Massey in [2] in the context of functions by "*each digits of the plaintext should influence many digits of the ciphertext*". In this section, since most cryptographic functions are boolean functions, we first formalize the interpretation of Massey and then we present some concepts quite close to our definition of diffusion in boolean functions.

We begin by introducing some notations and basic definitions. Let $\mathbb{F}_2$ denote the Galois field of two distinguished elements (denoted by 0 and 1). Let p and q be two non-negative natural numbers. A *boolean function* is a function $F : \mathbb{F}_2^p \longrightarrow \mathbb{F}_2$. A *generalized boolean function* is a function $F : \mathbb{F}_2^p \longrightarrow \mathbb{F}_2^q$. We denote by $F_j : \mathbb{F}_2^p \longrightarrow \mathbb{F}_2$ the j^{th} *component function* of a generalized boolean function F. We denote by d_H the Hamming distance of any vector space on $\mathbb{F}_2$. In order to define the diffusion for boolean functions, we introduce for all $i \in \{1, \ldots, p\}$ the following equivalence relation on $\mathbb{F}_2^p : x \approx_i y$ if $x_k = y_k$ for almost all $k \in \{1,\ \ldots\ p\} - \{i\}$, *i.e.* x and y can only differ on their i^{th} bit. There are exactly 2^{p-1} equivalence classes. The set of equivalence classes is denoted by $\mathcal{C}l_i$.

Each equivalence class $Q \in \mathcal{C}l_i$ contain exactly two elements of $\mathbb{F}_2^p$ denoted by Q^0 and Q^1 such that $Q^0 = 0$ and $Q^1 = 1$.

Definition 1. *Let $F : \mathbb{F}_2^p \longrightarrow \mathbb{F}_2^q$. The diffusion of F according to a coordinate $i \in \{1, \ldots, p\}$ and an equivalence class $Q \in \mathcal{C}l_i$ is $diff_F(i,Q) = d_H(F(Q^0), F(Q^1))$.*

It represents the number of output bits which changes when the i^{th} bit of a particular input is complemented. The *minimum diffusion* along the coordinate i is defined by $diff_{\min_F}(i) = \min_{Q \in \mathcal{C}l_i}(diff_F(i,Q))$ and the *average diffusion* along i is $diff_{avg_F}(i) = \frac{1}{2^{p-1}} \sum_{Q \in \mathcal{C}l_i} diff_F(i,Q)$.
We introduce the following definition, in order to take in account all the coordinates :

Definition 2. *Let $F : \mathbb{F}_2^p \longrightarrow \mathbb{F}_2^q$. The minimum diffusion of F is defined by $diff_{\min_F} = \min_{i \in \{1,\ldots,p\}} diff_{\min_F}(i)$.*
The average diffusion of F is defined by $diff_{avg_F} = \frac{1}{p} \sum_{i=1}^{p} diff_{avg_F}(i)$.

Example 1. Let $\oplus$ denote the XOR operation on $\mathbb{F}_2$. Let the one-time pad function $OTP_\oplus : (\mathbb{F}_2^p)^2 \longrightarrow \mathbb{F}_2^p$ defined by $OTP_\oplus(x,y) = (x_1 \oplus y_1, \ldots, x_p \oplus y_p)$. We have $diff_{min_{OTP_\oplus}} = 1$.

Example 2. Let C be a linear code on $\mathbb{F}_2$ with length n, dimension k and minimum distance d. If G denotes its generating matrix, we have for $g : \mathbb{F}_2^k \longrightarrow \mathbb{F}_2^n$ defined by $g(x) = x.G$, $diff_{min_g} = d$.

Some authors such as Massey use another definition for diffusion : the *diffusion on symbols*. Let $F : \mathbb{F}_2^{n.p} \longrightarrow \mathbb{F}_2^{m.q}$ where n and m are two non-negative integers called respectively the size of input symbols and the size of output symbols. An input symbol is thus an element of $\mathbb{F}_2^n$, an output symbol is an element of $\mathbb{F}_2^m$ and F can be seen as a function from p input symbols to q input symbols. If $X \in \mathbb{F}_2^{n.p}$, we denote by X_i (for $i \in \{1, \ldots, p\}$) the i^{th} symbol of X. For $j \in \{1, \ldots, q\}$ and $(X_1, \ldots, X_p) \in (\mathbb{F}_2^n)^p$, we denote by F_j the j^{th} *symbol component function* of F defined by

$$\begin{array}{l} F_j : (\mathbb{F}_2^n)^p \longrightarrow \mathbb{F}_2^m \\ (X_1, \ldots, X_p) \mapsto (\Pi_{(j-1)m+1}(F(X_1, \ldots, X_p)), \ldots, \Pi_{jm}(F(X_1, \ldots, X_p))) \end{array} \quad (1)$$

where Π_k denotes the k^{th} projection from $\mathbb{F}_2^{m.q}$ to $\mathbb{F}_2$.
We also define for $(Y_1, Y_2) \in (\mathbb{F}_2^m)^2$,

$$1_H(Y_1, Y_2) = \begin{cases} 1 \; if \quad d_H(Y_1, Y_2) \geq 1 \\ 0 \; else \end{cases} . \quad (2)$$

So we can define the diffusion on symbols by the minimum number of output symbols which vary when we complement one bit of the i^{th} ($i \in \{1, \ldots, p\}$) input symbol X_i of a particular $X = (X_1, \ldots, X_p) \in (\mathbb{F}_2^n)^p$:

$$\begin{array}{l} diff_F(i,X) = \min_{Z \in C_{\mathbb{F}_2^n}(X_i,1)} \sum_{j=1}^{q} 1_H(F_j(X), \\ \qquad\qquad F_j(X_1, \ldots, X_{i-1}, Z, X_{i+1}, \ldots, X_p)) \end{array} \quad (3)$$

where $C_{\mathbb{F}_2^n}(X_i, 1)$ denotes the unite circle centered on X_i in $\mathbb{F}_2^n$ with the Hamming distance. Finally we define the minimum diffusion of F by

$$diff_{\min_F} = \min_{(i,X)\in\{1,\ \ldots,\ p\}\times F_2^n} diff_F(i, X) \ . \tag{4}$$

Note that this notion of diffusion on symbols coincides with the previous one on bits if $n = m = 1$. Even if the diffusion on symbols is a strict generalization of diffusion of bits, is seems to be a less accurate measure of the diffusion of information.

We have introduced the natural definitions of the concept of diffusion. Some concepts which are close to the diffusion one are exposed in the following. First of all, a function $F : \mathbb{F}_2^p \longrightarrow \mathbb{F}_2^q$ is *complete* if for all $i \in \{1,\ \ldots,\ p\}$ and for all $j \in \{1,\ \ldots,\ q\}$, there exists $Q_{(i,j)} \in \mathcal{C}l_i$ such that $d_H(F_j(Q^0_{(i,j)}), F_j(Q^1_{(i,j)})) \geq 1$. In other terms, all output bits of F is dependant of each input bits. The completeness has been introduced by Kam and Davida in [3].
Another important concept is the *avalanche effect* which means, if exhibited by a boolean function, that an average of half the number of output bits varies when only one input bit is complemented. In order to take in account the two previous notions, Webster and Tavares [4] have introduced the *strict avalanche criteria* or *S.A.C.* : if a function satisfies *S.A.C.*, every output bits change with a probability of $\frac{1}{2}$ every times only one input bit is complemented. Clearly, a function which satisfies *S.A.C.* is complete and exhibits the avalanche effect.

These last characteristics are very restrictive since they should be verified only by functions which provide a great diffusion on bits. In this sense, the notion of diffusion introduced first, seems to be a more primitive and relevant measure for the diffusion or propagation of an input bit of information along the output bits but suffers from a lack of description for the localization of the output bits which have changed.
In the next section, we present the *complete diffusion* property of boolean functions introduced by Massey in his design of IDEA. As one will see this notion is quite close to a completeness on symbols instead of bits.

4 Complete diffusion and computational graphs

In [5], Massey used abundantly computational graphs to represent boolean functions of IDEA cryptosystem and also for the diffusion part of its design. It should be interesting to present the computational graphs in a general way and then to describe their use in the particular context of diffusion with respect to Massey's concept.

Let F be a generalized boolean function of the following type :

$$\begin{array}{lll} F : \mathbb{F}_2^{n_1} \times \ldots \times \mathbb{F}_2^{n_p} & \longrightarrow & \mathbb{F}_2^{m_1} \times \ldots \times \mathbb{F}_2^{m_q} \\ (X_1,\ \ldots,\ X_p) & \mapsto & (F_1(X_1,\ \ldots,\ X_p),\ \ldots,\ F_q(X_1,\ \ldots,\ X_p)) \end{array} \tag{5}$$

with for $i \in \{1, \ldots, p\}$, $X_i \in \mathbb{F}_2^{n_i}$ and for all $j \in \{1, \ldots, q\}$, $F_j(X_1, \ldots, X_p) = (\Pi_{m_{j-1}+1}(F(X_1, \ldots, X_p)), \ldots, \Pi_{m_{j-1}+m_j}(F(X_1, \ldots, X_p))) \in \mathbb{F}_2^{m_j}$ (with $m_0 = 0$ by convention).
Each F_j is composed of several functions, each with less input symbols. Thus $F_j(X_1, \ldots, X_p)$ is a term t of a formal language with the (free) variables in $\{X_1, \ldots, X_p\}$ and which contains functional sub-terms. Let F be a boolean function. For a specific decomposition $F(X_1, \ldots, X_p) = (t_1, \ldots, t_q)$ (with each X_i occurring in at less one t_j), we can construct an *acyclic directed graph* with labeled nodes and edges, called *computational graph* of the decomposition $(t_1, \ldots, t_q)$, as follows :

1. We construct the q trees of the terms t_1 to t_q. The nodes are labeled by variable names (the *input variables*) or operation names, the edges are oriented in the direction argument to function and labeled by the position of the argument in the function.
2. For each $j \in \{1, \ldots, q\}$, we add for the tree of t_j a node labeled by Y_j (if $j_1 \neq j_2$ then $Y_{j_1} \neq Y_{j_2}$) and an edge from the root of the tree to this new node. The Y_j are the *output variables*.
3. We identify the common sub-trees (we obtain then an acyclic graph).

Example 3. A computational graph of $F(X_1, X_2, X_3) = (f_2(f_1(X_1), X_3), f_4(f_3(X_2), f_2(f_1(X_1), X_3)))$ is displayed in Fig. 1.

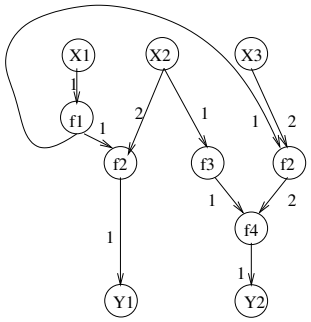

Fig. 1. Computational Graph of $F(X_1, X_2, X_3)$

We can show that all generalized functions has at less one binary computational graph, *i.e.* the operations which names occur in the graph are binary.
Now we can introduce the complete diffusion of Massey and then relate it with computational graphs. Let F be a generalized function of the form (5), $X = (X_1, \ldots, X_p)$ a fixed vector of p symbols and $i \in \{1, \ldots, p\}$ fixed. We define the set of vectors of p symbols $Z = (Z_1, \ldots, Z_p)$ which differ of X only on the

i^{th} symbol by :

$$\mathscr{D}(X,i) = \{Z \in \mathbb{F}_2^{n_1} \times \ldots \times \mathbb{F}_2^{n_p} | \forall k \in \{1, \ldots, p\} - \{i\}, Z_k = X_k \text{ and } Z_i \neq X_i\} \ . \tag{6}$$

Definition 3. *A function $F : \mathbb{F}_2^{n_1} \times \ldots \times \mathbb{F}_2^{n_p} \longrightarrow \mathbb{F}_2^{m_1} \times \ldots \times \mathbb{F}_2^{m_q}$ is said to have complete diffusion (on symbols) if $\forall i \in \{1, \ldots, p\}$ and $\forall j \in \{1, \ldots, q\}$ there exit $X \in \mathbb{F}_2^{n_1} \times \ldots \times \mathbb{F}_2^{n_p}$ and $Z \in \mathscr{D}(X,i)$ such that $d_H(F_j(X), F_j(Z)) \geq 1$.*

In other terms, each output symbols depends of every input symbols. This definition is less restrictive than the completeness property presented in the previous section but is quite similar.
Massey used this notion on particular functions : the (Massey's) *cipher functions*. They are defined as follows : a function $F : (\mathbb{F}_2^n)^2 \times (\mathbb{F}_2^k)^2 \longrightarrow (\mathbb{F}_2^m)^2$ is a *cipher function* if for all fixed $(Z_1, Z_2) \in (\mathbb{F}_2^k)^2$, the function $(X_1, X_2) \in (\mathbb{F}_2^n)^2 \mapsto F(X_1, X_2, Z_1, Z_2) \in (\mathbb{F}_2^m)^2$ is invertible. The function called *multiplication-addition* function defined in IDEA is a cipher function with complete diffusion.
The theorem which binds the notion of complete diffusion and computational graphs is the following one :

Theorem 1. *If a cipher function F has complete diffusion, then any binary computational graph of F contains at least four operations.*

This theorem allows quick refutations of complete diffusion of functions just by checking the number of operation nodes of a binary computational graph.

We add a weak necessary condition for complete diffusion of generalized boolean functions : if F has complete diffusion then any computational graph of F completes *symbolic diffusion i.e.*, for each couple of nodes (n_O, n_I) in the graph labeled with an output variable (for n_O) and with an input variable (for n_I), there exists an oriented path in the graph from n_I to n_O.
The proof is easy. Suppose by contradiction that there exists (n_O, n_I), n_O labeled by Y_j and n_I labeled by X_i such that there is no path in the graph from n_I to n_O. Then by construction of the graph, X_i does not occur in the term t_j which represents $F_j(X_1, \ldots, X_p)$ and thus F_j does not depend of X_i which is a contradiction with the complete diffusion of F. Note that the reciprocal is false. For instance, let $F(X_1, X_2) = F_2(F_1(X_1, X_2), F_2(X_1, X_2))$ with

$$F_1(X) = \begin{cases} 1 \text{ if } & X = (1,1) \\ 0 \text{ else} & \end{cases} \qquad F_2(X) = \begin{cases} 1 \text{ if } & X \neq (1,1) \\ 0 \text{ else} & \end{cases} \ . \tag{7}$$

Then $F(X_1, X_2) = 0$ for all (X_1, X_2) and thus F does not have complete diffusion (whereas both F_1 and F_2 have complete diffusion). But we can see that, if we denote the output variable of F by Y, then for all computational graph of F there exists a path from X_1 to Y and from X_2 to Y.

The complete diffusion of Massey provides a good qualitative measure for the diffusion because it is not too restrictive and not trivial anymore. The constraints on functions to provide the complete diffusion are less strong than the

ones for the completeness of bits or the avalanche effect. However it does not give a quantitative information about the diffusion of symbols along boolean functions : it can be regarded as a fundamental design principle to construct functions with good diffusion but not as a goal to accomplish. Indeed, complete diffusion on symbols should be one of the expected properties of high-diffusion functions : we want the high diffusion of information not to be localized in a little part of the results because we keep in mind the historical notion of diffusion.
We can add that this use of graph in diffusion of information problems has been continue by Massey to define the optimal diffusion of SAFER+ (see [6]).

In the last section, we describe the point of view which was followed by Daemen and Rijmen to conceive the AES in order to be resistant against differential and linear cryptanalysis.

5 Diffusion in Rijndael

The diffusion requirement in [9] of Rijndael or AES cryptosystem is explicitly designed to provide good resistance against the so-called linear and differential cryptanalysis.
Differential cryptanalysis has been described by Biham and Shamir in [7]. It is a chosen-plaintext attack of which the goal is to determine the key used to encrypt messages. The principle of this method is, with a fixed choice of two plaintexts, to study the evolution of the differences between the two intermediate ciphertexts at each round encrypted by the same (and unknown) key. By analyzing final differences of pairs of ciphertexts, we deduce several probabilities for keys and thus the most probable key.
Linear cryptanalysis was presented first by Matsui in [8]. It is a known plaintext attack. The fundamental idea of this attack is to approximate a (non-linear) block-cipher with linear expressions of several bits of plaintext and bits of intermediate ciphertexts in order to find bits of key (or sub-key).
As diffusion notion in Rijndael is based on these two attacks, we need to introduce some mathematical objects used in differential and linear cryptanalysis. A *parity* of a given boolean vector is a boolean function that consists of the XOR $\oplus$ of a number of bits. A parity is determined by the bit positions of the boolean vector that are included in the XOR. The *selection pattern* V of a parity is a boolean vector that has 1 in the components that are included in the parity and 0 in all others. We express the parity of vector X according to the selection pattern V by $V^\intercal.X$ in a matrix format. For instance, let V be the selection pattern $(0,1,0,1,1)$ in $\mathbb{F}_2^5$ and $X = (X_1, X_2, X_3, X_4, X_5)$ be any vector of $\mathbb{F}_2^5$. The parity of X according to the selection pattern V is then $X_2 \oplus X_4 \oplus X_5$. The parities occur in linear cryptanalysis as linear expressions. Moreover this attack exploits correlations between two boolean functions f and g, defined by $C(f(X), g(X)) = 2.Prob(f(X) = g(X)) - 1$ where $Prob$ stands for a given measure of probability. If F denotes a generalized boolean function from $\mathbb{F}_2^p$ to $\mathbb{F}_2^q$, we define the $2^q \times 2^q$ correlation matrix $C^{(F)}$ of correlation between input and

output parities of F by defining the element $C_{U,V}^{(F)}$ in row U and column V equals to $C(U^\intercal F(X), V^\intercal X)$. In differential cryptanalysis the basic objects are the *difference patterns* : there are bit-wise XOR of two vectors $X \oplus Y$.
The principle of diffusion in Rijndael is based on the quantitative study of correlation and differential propagations along a generalized boolean function (the *round transformation*) $\rho : (\mathbb{F}_2^n)^p \longrightarrow (\mathbb{F}_2^n)^p$. ρ is (essentially, because we do not take into account the XOR with sub-keys) the composition of two generalized boolean functions γ and then λ, where γ is non-linear transformation which applies on each of the p symbols (of n bits) independently and λ is a linear transform. In the context of substitution-permutation networks, γ should be identified with an S-Box and λ with a P(ermutation)-Box. λ mixes in linear expressions the output symbols of γ thus the power of diffusion is essentially localized in λ. The fundamental results of diffusion in Rijndael are given in a simple two-round model. We define the *active symbols* or *weight* of a (differential or selection) pattern X its non-zero symbols and it is denoted by $w(X)$. A two-round (differential or linear) *trail* is a couple $(X, \rho(X))$ where X is a (difference or selection) pattern. We define then the *active symbols* of a two-round trail as $w(X) + w(\rho(X))$. The main idea of diffusion in Rijndael (in a two-round context) is to define the diffusion by the number of active symbols in two-round trails and to describe mechanisms which eliminate low-weight trails. Indeed, the higher are the weights of trails, the harder are the respective cryptanalysis. A relevant measure of diffusion in this context is the minimum number of active symbols at the input and the output of ρ. It is called the *branch number* of ρ. It gives a lower bound for the propagation of differences and linear expressions, which is more or less the notion of diffusion chosen by Daemen and Rijmen. Formally, the differential branch number of a transformation ρ is given by $\mathscr{B}_d(\rho) = \min_{X,Y \neq Y}\{w(X \oplus Y) + w(\rho(X) \oplus \rho(Y))\}$ and its linear branch number, by $\mathscr{B}_l(\rho) = \min_{U,V,C(U^\intercal X, V^\intercal \rho(X)) \neq 0}\{w(U) + w(V)\}$. The fundamental theorem in this context is then :

Theorem 2. *For a block-cipher with round function $\rho = \lambda \circ \gamma$ of the previous type, the number of active symbols of any two-round trail is lower bounded by the branch number of λ.*

This is generalized at a multi-round level in [9]. This approach of diffusion is to be relied with works of Chabaud and Vaudenay [10] which show that, in some cases, differential-resistant and linear-resistant functions are essentially the same. It shows that there exist underlying links between the three solidity criteria quoted in introduction and it could be an interesting research direction to establish a general theory of diffusion of information. Moreover, without thinking in terms of resistance against statistical attacks, this notion of diffusion mixes the quantitative effect of propagation of a certain type of information and the localization of this effect which are two fundamental axes of research in the problems of diffusion.

6 Summary and Conclusion

We have presented four notions of diffusion during this paper. The first one, the *method of diffusion* introduced by Shannon, can be used to delimit in an abstract way the area of the problems of diffusion but is not very useful to build high-diffusion functions. Then we have seen a quantitative characterization of diffusion at the bit level which seems to be a relevant measure of diffusion associated with other concepts such as completeness. The two last concepts were the *complete diffusion* by Massey and the *branch number* of Daemen and Rijmen and in both approaches the granularity of the diffusion is at the symbol level and so less accurate but also less complex.
Finally we can note that all these notions study the links between a certain type of inputs (for instance, a change of one bit or symbols involved in linear expressions) and the corresponding output. It suggests that the diffusion concept should be a mathematical formalization of this underlying common point.

References

[1] C.E. Shannon : Communication Theory of Secrecy Systems. Bell Systems Technical Journal, Vol. 28, pp. 656-715, Oct. 1949

[2] J.L. Massey : An Introduction to Contemporary Cryptology. *Proc. IEEE*, Vol. 76, pp. 533-549, May 1988

[3] J.B. Kam, G.I. Davida : Structured Design of Substitution Permutation Encryption Networks. IEEE Transactions on Computers, Vol. 28, No. 10, 747, 1979

[4] A.F. Webster, S.E. Tavares : On the design of S-Boxes. *Advances in Cryptology, Proc. Crypto'85*, 1985

[5] X. Lai, J.L. Massey : A Proposal for a New Block Encryption Standard. *Advances in Cryptology, Proc. Eurocrypt'90*, pp. 389-404

[6] J.L. Massey : On the Optimality of SAFER+ Diffusion. *Second Advanced Encryption Standard Candidate Conference (AES2)*, Rome, Italy, on line available at http://csrc.nist.gov/encryption/aes/round1/conf2/aes2conf.htm

[7] E. Biham, A. Shamir : Differential Cryptanalysis of DES-like Cryptosystems. *Journal of Cryptology*, Vol. 4, No. 1, pp. 3-72, 1991

[8] M. Matsui : Linear Cryptanalysis Method for DES Cipher. *Advances in Cryptology, Proc. Eurocrypt'93, LNCS 809*, pp. 1-17, 1994

[9] J. Daemen, V. Rijmen : The Design of Rijndael. AES - The Advanced Encryption Standard. Ed. Springer-Verlag, Berlin, 2002

[10] F. Chabaud, S. Vaudenay : Links between Differential and Linear Cryptanalysis. Lecture Notes in Computer Science, Vol. 950, pp. 356-365, 1995