Approximations of a combining function and parity check equations

Anne Canteaut and Maria Naya-Plasencia

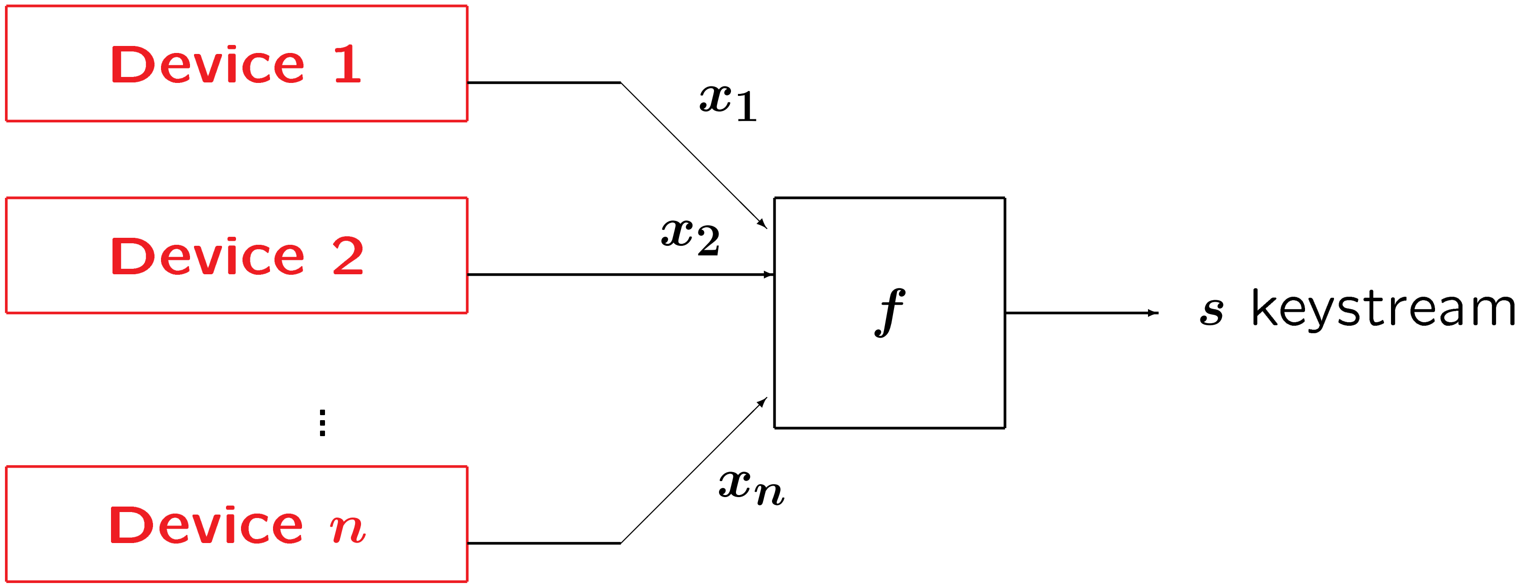
INRIA Paris-Rocquencourt
SECRET team (SEcurité, CRyptologie Et Transmissions)
Domaine de Voluceau
78153 Le Chesnay - France

Journées C2 2008

Outline

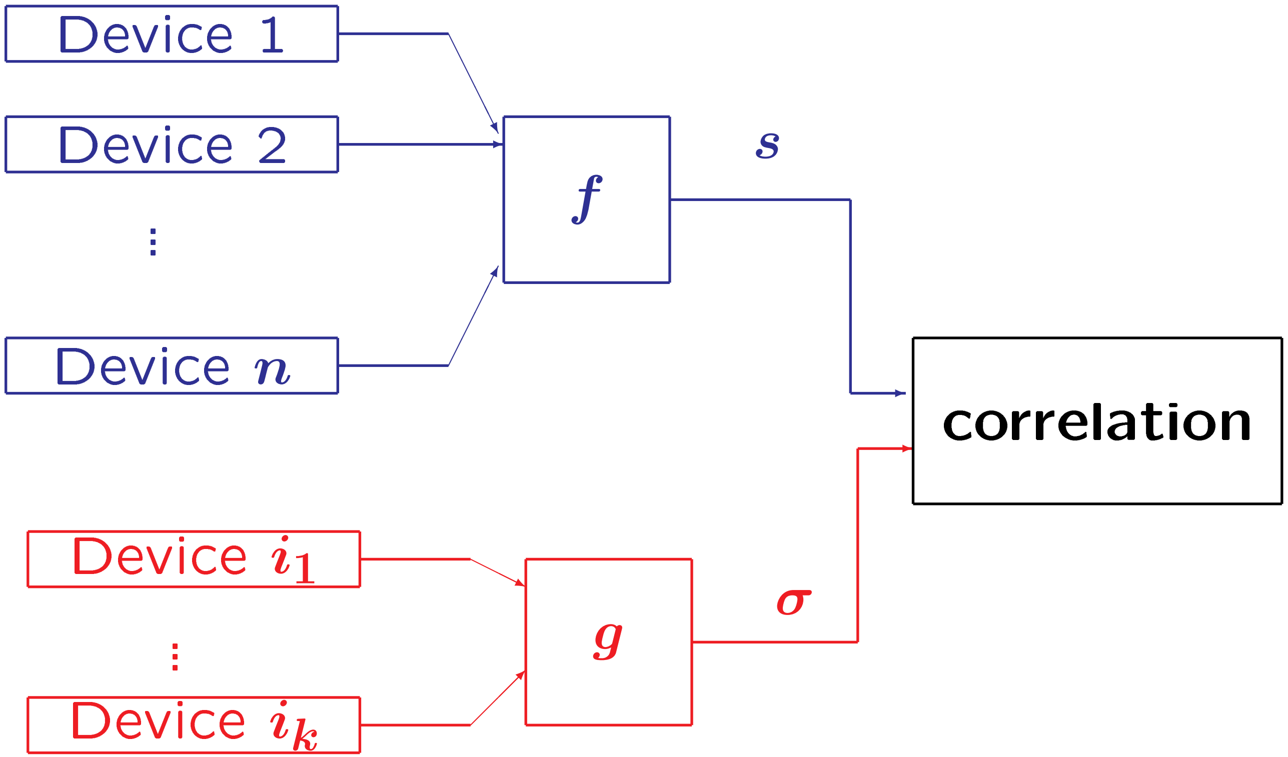
1. Divide-and-conquer attacks against some stream ciphers
2. Some attacks against Achterbahn-80
3. On the bias of parity check equations
4. Resilient functions

Combination generators for additive stream ciphers



where each x_i has period T_i.

Divide-and-conquer attack involving k constituent devices



where $\Pr[f(X_1, \ldots, X_n) = g(X_{i_1}, \ldots, X_{i_k})] > \frac{1}{2}$.

Resilient functions

Definition A Boolean function f is t-resilient if

$$\Pr[f(X_1, \ldots, X_n) = g(X_{i_1}, \ldots, X_{i_k})] = \frac{1}{2}$$

for any $k \leq t$ and for any function g of k variables.

The order of resiliency is the highest t such that f is t-resilient.

$\Longrightarrow$ we have to consider $t+1$ devices together.

Building parity-check relations [Johansson-Meier-Muller 06]

Property 1. $x_1 x_2 \ldots x_s$ has period $T_1 T_2 \ldots T_s$.

Property 2. Let $\sigma(t) = \sum_{i=1}^{s} x_i$ and

$$\mathcal{T} = \left\{ \sum_{i=1}^{s} c_i T_i, \;\; c_i \in \{0,1\} \right\}.$$

Then, for any $t \geq 0$,

$$\sum_{\tau \in \mathcal{T}} \sigma(t+\tau) = 0.$$

Example. For $\sigma = x_1 + x_2$:

$$\sigma(t) + \sigma(t+T_1) + \sigma(t+T_2) + \sigma(t+T_1+T_2) = 0, \;\; \forall t \geq 0.$$

Building parity-check relations [Johansson-Meier-Muller 06]

Let $\boldsymbol{\sigma} = \boldsymbol{g}(\boldsymbol{x_{i_1}}, \ldots, \boldsymbol{x_{i_k}})$.

For $\boldsymbol{g} = \sum_{i=1}^{m} \boldsymbol{m_i}(\boldsymbol{x_{i_1}}, ..., \boldsymbol{x_{i_k}})$, let us consider

$$\mathcal{T} = \left\{ \sum_{i=1}^{m} c_i m_i(T_{i_1}, ..., T_{i_k}), \quad c_i \in \{0, 1\} \right\}.$$

Then,

$$\sum_{\tau \in \mathcal{T}} \sigma(t + \tau) = 0.$$

Distinguishing attack [Johansson-Meier-Muller 06]

Let $s = f(x_1, \ldots, x_n)$ where

$$\Pr[f(X_1, \ldots, X_n) = g(X_{i_1}, \ldots, X_{i_k})] = \frac{1}{2}(1+\varepsilon) \text{ with } \varepsilon > 0.$$

For $g = \sum_{i=1}^{m} m_i(x_{i_1}, ..., x_{i_k})$ and

$$\mathcal{T} = \left\{ \sum_{i=1}^{m} c_i m_i(T_{i_1}, ..., T_{i_k}), \quad c_i \in \{0,1\} \right\}.$$

Then,

$$\Pr\left[\sum_{\tau \in \mathcal{T}} s(t+\tau) = 0 \right] \geq \frac{1}{2}(1+\varepsilon^{2^m}).$$

Complexity:

Time complexity $\simeq \varepsilon^{-2^{m+1}} \times 2^m$

Data complexity $\simeq \varepsilon^{-2^{m+1}} + g(T_{i_1}, \ldots, T_{i_k})$

Decimation by the period of a sequence [Hell-Johansson 06]

For $g = x_{i_j} + \sum_{i=1}^{m'} m_i(x_{i_1}, ..., x_{i_k})$, let us consider

$$\mathcal{T}' = \left\{ \sum_{i=1}^{m'} c_i m_i(T_{i_1}, ..., T_{i_k}), \;\; c_i \in \{0,1\} \right\}.$$

Then,

$$\Pr[\sum_{\tau \in \mathcal{T}'} s(t+\tau) = \sum_{\tau \in \mathcal{T}'} x_{i_j}(t+\tau)] \geq \frac{1}{2}(1+\varepsilon^{2^{m'}}),$$

implying

$$\Pr\left[\sum_{\tau \in \mathcal{T}'} s(tT_{i_j} + \tau) = \text{ cst}\right] \geq \frac{1}{2}(1+\varepsilon^{2^{m'}}),$$

Complexity:

Time complexity $\simeq \varepsilon^{-2^{m'+1}} \times 2^{m'}$

Data complexity $\simeq \varepsilon^{-2^{m'+1}} T_{i_j} + g'(T_{i_1}, \ldots, T_{i_k})$

Initial state recovery [Johansson-Meier-Muller 06]

For $g = \sum_{j=1}^{s} x_{i_j} + \sum_{i=1}^{m'} m_i(x_{i_1}, ..., x_{i_k})$, let us consider

$$\mathcal{T}' = \left\{ \sum_{i=1}^{m'} c_i m_i(T_{i_1}, ..., T_{i_k}), \quad c_i \in \{0,1\} \right\}.$$

Then,

$$\Pr[\sum_{\tau \in \mathcal{T}'} s(t+\tau) + \sum_{j=1}^{s} \sum_{\tau \in \mathcal{T}'} x_{i_j}(t+\tau) = 0] \geq \frac{1}{2}(1+\varepsilon^{2^{m'}}).$$

Attack:

Perform an exhaustive search for the initial states of Dev $i_1, \ldots, i_s$.
For each possible initial state, compute the parity-check equations.

Complexity:

Data complexity $\simeq \varepsilon^{-2^{m'+1}} 2 \ln 2 (L_{i_1} + \ldots + L_{i_s}) + g'(T_{i_1}, \ldots, T_{i_k})$

Time complexity $\simeq \varepsilon^{-2^{m'+1}} 2 \ln 2 (L_{i_1} + \ldots + L_{i_s}) \times 2^{m'} \times 2^{L_{i_1} + \ldots + L_{i_s}}$

Achterbahn-80 [Gammel-Göttfert-Kniffler06]

11 NLFSRs of length $L_i = 21 + i$ and of period $T_i = 2^{L_i} - 1$, $1 \leq i \leq 11$.

f: 6-resilient combining function of degree 4:

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_7 + x_9 + x_{11} + x_2x_{10} + x_2x_{11} + x_4x_8 + x_5x_6 + x_6x_8 + x_6x_{10} + x_6x_{11} + x_7x_8 + x_8x_9 + x_8x_{10} + x_9x_{10} + x_9x_{11} + x_1x_2x_8 + x_1x_4x_{10} + x_1x_4x_{11} + x_1x_8x_9 + x_1x_9x_{10} + x_1x_9x_{11} + x_2x_3x_8 + x_2x_4x_8 + x_2x_4x_{10} + x_2x_4x_{11} + x_2x_7x_8 + x_2x_8x_{10} + x_2x_8x_{11} + x_2x_9x_{10} + x_2x_9x_{11} + x_3x_4x_8 + x_3x_8x_9 + x_4x_7x_8 + x_4x_8x_9 + x_5x_6x_8 + x_5x_6x_{10} + x_5x_6x_{11} + x_6x_8x_{10} + x_6x_8x_{11} + x_7x_8x_9 + x_8x_9x_{10} + x_8x_9x_{11} + x_1x_2x_3x_8 + x_1x_2x_7x_8 + x_1x_3x_5x_8 + x_1x_3x_8x_9 + x_1x_4x_8x_{10} + x_1x_4x_8x_{11} + x_1x_5x_7x_8 + x_1x_7x_8x_9 + x_1x_8x_9x_{10} + x_1x_8x_9x_{11} + x_2x_3x_4x_8 + x_2x_3x_5x_8 + x_2x_4x_7x_8 + x_2x_4x_8x_{10} + x_2x_4x_8x_{11} + x_2x_5x_7x_8 + x_2x_8x_9x_{10} + x_2x_8x_9x_{11} + x_3x_4x_8x_9 + x_4x_7x_8x_9 + x_5x_6x_8x_{10} + x_5x_6x_8x_{11}$$

First attack against Achterbahn-80

Quadratic approximation:

$$x_1 + x_2 + x_7 + x_3x_{10} + x_4x_9, \quad \varepsilon = 2^{-5}.$$

$$\mathcal{T} = \{c_1T_3T_{10} + c_2T_4T_9, \;\; c_1, c_2 \in \{0, 1\}\}$$

- Decimation by T_7
- Exhaustive search on R1 and R2.

For $\sigma = x_1 + x_2$,

$$s(tT_7) + s(tT_7 + T_3T_{10}) + s(tT_7 + T_4T_9) + s(tT_7 + T_3T_{10} + T_4T_9) =$$
$$\sigma(tT_7) + \sigma(tT_7 + T_3T_{10}) + \sigma(tT_7 + T_4T_9) + \sigma(tT_7 + T_3T_{10} + T_4T_9) + \text{cst}$$

with bias $\geq 2^{-20}$.

Data complexity $= 2^{74}$ Time complexity $= 2^{91}$.

First attack against Achterbahn-80 [Hell-Johansson06]

The exact bias of

$$s(tT_7) + s(tT_7 + T_3T_{10}) + s(tT_7 + T_4T_9) + s(tT_7 + T_3T_{10} + T_4T_9) =$$
$$\sigma(tT_7) + \sigma(tT_7 + T_3T_{10}) + \sigma(tT_7 + T_4T_9) + \sigma(tT_7 + T_3T_{10} + T_4T_9) + \text{cst}$$

is not 2^{-20} but 2^{-12}.

Then,

Data complexity $= 2^{58.3}$ Time complexity $= 2^{75}$.

Second attack against Achterbahn-80 [Naya-Plasencia06]

Linear approximation:

$$x_1 + x_2 + x_7 + (x_3 + x_{10}) + (x_4 + x_9), \quad \varepsilon = 2^{-3}.$$

$$\mathcal{T} = \{c_1 T_3 T_{10} + c_2 T_4 T_9, \quad c_1, c_2 \in \{0,1\}\}$$

- Decimation by T_7
- Exhaustive search on R1 and R2.

For $\sigma = x_1 + x_2$,

$$s(tT_7) + s(tT_7 + T_3T_{10}) + s(tT_7 + T_4T_9) + s(tT_7 + T_3T_{10} + T_4T_9) =$$
$$\sigma(tT_7) + \sigma(tT_7 + T_3T_{10}) + \sigma(tT_7 + T_4T_9) + \sigma(tT_7 + T_3T_{10} + T_4T_9) + \text{cst}$$

with bias $\geq 2^{-12}$.

Data complexity $= 2^{58.3}$ Time complexity $= 2^{75}$.

Related issues

- Is the exact bias always given by the bias of the linear approximation?
- Can we get a better result with higher degree approximations?
- Can we build better parity checks (higher bias) from approximations with more than t+1 variables?

Reminder on parity-check relations

$$\begin{aligned} h(x_1,\ldots,x_n) &= f(x_1,\ldots,x_n) + g(x_{j_1},\ldots,x_{j_{s+k}}) \\ &= f'(x_1,\ldots,x_n) + g'(x_{j_1},\ldots,x_{j_s}) \end{aligned}$$

has bias ε.

$$pc(t) = \sum_{\tau\in\mathcal{T}} h(t+\tau) = \sum_{\tau\in\mathcal{T}} f'(t+\tau).$$

$$\Pr[pc(t) = 0] \geq \frac{1}{2}(1+\varepsilon^{2^m}).$$

Examples on building parity-check relations

- $g_1(x_1, x_2, x_3, x_4, x_7, x_9, x_{10}) = x_1 + x_2 + x_7 + x_3x_{10} + x_4x_9$
  $h_1(x_1, \ldots, x_{11}) = f'(x_1, \ldots, x_{11}) + x_3x_{10} + x_4x_9$. $\varepsilon = 2^{-5}$.
- $g_2(x_1, x_2, x_3, x_4, x_7, x_9, x_{10}) = x_1 + x_2 + x_7 + x_3 + x_{10} + x_4 + x_9$
  $h_2(x_1, \ldots, x_{11}) = f'(x_1, \ldots, x_{11}) + x_3 + x_{10} + x_4 + x_9$. $\varepsilon = 2^{-3}$.

$$\begin{aligned} pc(t) &= h_i(t) + h_i(t + T_3T_{10}) + h_i(t + T_4T_9) + h_i(t + T_3T_{10} + T_4T_9) \\ &= f'(t) + f'(t + T_3T_{10}) + f'(t + T_4T_9) + f'(t + T_3T_{10} + T_4T_9), \end{aligned}$$

$\varepsilon = 2^{-12}$.

Building parity-check relations from one approximation

- $g_2(x_1, x_2, x_3, x_4, x_7, x_9, x_{10}) = x_1 + x_2 + x_7 + x_3 + x_{10} + x_4 + x_9$
  $h_{2_i}(x_1, \ldots, x_{11}) = f'_i(x_1, \ldots, x_{11}) + g'_j(x_{j_1}, \ldots, x_{j_s})$. $\varepsilon = 2^{-3}$.

  $pc_1(t) = f'_1(t) + f'_1(t + T_3T_{10}) + f'_1(t + T_4T_9) + f'_1(t + T_3T_{10} + T_4T_9)$

  $\varepsilon \geq 2^{-12}$.

  $pc_2(t) = f'_1(t) + f'_1(t + T_3) + f'_1(t + T_4T_{10}T_9) + f'_1(t + T_3 + T_4T_{10}T_9)$

  $\varepsilon \geq 2^{-12}$.

  $pc_3(t) = f'_3(t) + f'_3(t + T_1T_2T_7) + f'_3(t + T_3T_4T_9T_{10}) + f'_3(t + T_1T_2T_7 + T_3T_4T_9T_{10})$

  $\varepsilon \geq 2^{-12}$.

- Any parity check can be generated by an affine approximation/function.

- What is the exact bias of each pc(t)?

Approximation of a resilient function

Theorem [Canteaut-Trabbia 00] [Zhang 00]
Let f be t-resilient function of n variables. Then, for any K of size $t+1$ the best approximation is achieved by the **affine function**

$$\sum_{i \in K} x_i + \varepsilon, \; \varepsilon \in \{0,1\} \; .$$

Bias of parity-checks involving $(t+1)$ variables

Theorem [Naya-Plasencia 07]
Let f be t-resilient function. The bias of any parity-check equation built from a $(t+1)$-variable linear approximation of f with bias ε is ε^M where M is the number of terms in the parity-check equation.

Examples of parity-checks involving $(t+1)$ variables

- $g_2(x_1, x_2, x_3, x_4, x_7, x_9, x_{10}) = x_1 + x_2 + x_7 + x_3 + x_{10} + x_4 + x_9$
  $h_{2_i}(x_1, \ldots, x_{11}) = f'_i(x_1, \ldots, x_{11}) + g'_j(x_{j_1}, \ldots, x_{j_s})$. $\varepsilon = 2^{-3}$.

  $pc_1(t) = f'_1(t) + f'_1(t + T_3T_{10}) + f'_1(t + T_4T_9) + f'_1(t + T_3T_{10} + T_4T_9)$
  $\varepsilon = 2^{-12}$.

  $pc_2(t) = f'_1(t) + f'_1(t + T_3) + f'_1(t + T_4T_{10}T_9) + f'_1(t + T_3 + T_4T_{10}T_9)$
  $\varepsilon = 2^{-12}$.

  $pc_3(t) = f'_3(t) + f'_3(t + T_1T_2T_7) + f'_3(t + T_3T_4T_9T_{10}) + f'_3(t + T_1T_2T_7 + T_3T_4T_9T_{10})$
  $\varepsilon = 2^{-12}$.

$$pc_4(t) = f'_4(t) + f'_4(t + T_1T_3T_7), \ \varepsilon = 2^{-6}.$$

What happens with $t+k$ variables when $k>1$?

- $\boldsymbol{f(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_1x_3}$. 0-resilient.
- We consider $\boldsymbol{g = x_1 + x_2}$, with $\varepsilon = 0$.
- We build the parity check associated to that function:

$$pc(t) = f(t) + f(t+T_1) + f(t+T_2) + f(t+T_1T_2)$$

- $\Pr[\boldsymbol{pc(t) = 0}] = \boldsymbol{\frac{1}{2}(1 + 2^{-3}) \neq \frac{1}{2}(1 + 0^4) = 0.}$

Bias of parity-checks involving $(t+k)$ variables

- $h'(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) + x_1 + \ldots + x_{t+1}$, with ε'.
- $h(x_1, \ldots, x_n) = h'(x_1, \ldots, x_n) + x_{t+2} + \ldots + x_{t+k}$, with ε.

$$pc(t) = \sum_{\tau_1 \in <T_{t+2}, \ldots, T_{t+k}>} \sum_{\tau_2 \in <T_1, \ldots, T_{t+1}>} f(t + \tau_1 + \tau_2)$$

$$\Pr[pc(t) = 0] \geq \frac{1}{2}(1 + {\varepsilon'}^{2^{t+k}}).$$

where $\varepsilon' = max_{\alpha, wt(\alpha)=t+1} \varepsilon(f + \alpha(x_1, \ldots, x_{t+k}))$

Previous example with $0+2$ variables

- With the 0-resilient function f:

$$f(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_1x_3.$$

  $\boldsymbol{x_1}$ is an approximation of f with bias $\varepsilon = 2^{-1}$.

- We consider the previously defined parity check, that can be derived from $\boldsymbol{g = x_1 + x_2,}$ that has a bias $\varepsilon_g = 0$.

$$pc(t) = f(t) + f(t + T_1) + f(t + T_2) + f(t + T_1T_2)$$

- $\mathrm{Pr}[\boldsymbol{pc(t) = 0}] = \boldsymbol{\frac{1}{2}(1 + 2^{-3}) \geq \frac{1}{2}(1 + (2^{-1})^{2^2} = \frac{1}{2}(1 + 2^{-4}).}$

- Question: is it possible that

  $\mathrm{Pr}[\boldsymbol{pc(t) = 0}] > \boldsymbol{\frac{1}{2}(1 + \varepsilon'^{2^M}})$?

Trade-off on divide-and-conquer attacks

- At the attacks of the type that we have described, to find the best complexity we have to make a trade-off between several parameters affecting time complexity and data complexity.

- With a combining function $f(x_1, \ldots, x_n)$ we can build a parity check equation with the highest possible bias, $\varepsilon = 1$, and with the lowest possible number of terms:

$$f(t) + f(t + T_1 T_2 \ldots T_n).$$

  This parity check equation needs $T_1 T_2 \ldots T_n$ bits of keystream to be computed. In the case of Achterbahn as in the case of all the other reasonable algorithms, this quantity is much too high.

Conclusions

So we have found some information about the bias of parity checks when using t-resilient combining functions, which will be the case in cryptographic applications as the one described.

For a t-resilient combining function:

- the bias of any parity-check relation involving $(t+1)$ variables is derived from the bias of the corresponding linear approximation.
- the bias of any parity-check relation involving more than $(t+1)$ variables has a lower bound that is the bias of its corresponding best linear approximation of $t+1$ variables raised to the number of terms of the parity check.