Analysis of pseudorandom sequences

Viktória Tóth
Eötvös Loránd University, Budapest, Hungary
Department of Computer Algebra

Summer School on Real-world Crypto and Privacy
June 5–9, 2017
Sibenik, Croatia

- Introduction
- New, constructive approach
- Collision and avalanche effect
  - definitions
  - analysis of constructions
- Results
- Conclusion

Pseudorandomness:
- numerical analysis, pure mathematics, cryptography
- keystream in Vernam-cipher:

$$m + k = c$$

New, constructive approach:
Mauduit and Sárközy, 1996

Advantages:

1. More constructive
2. No use unproved hypothesis
3. Describe the single sequences
4. Apriori testing
5. Characterizing with real-valued function

   $\Rightarrow$ comparableness

Measures

For a given sequence $E_N = (e_1, ..., e_N) \in \{-1, +1\}^N$ **the correlation measure of order** k **of** E_N **is:**

$$C_k(E_N) = \max_{M,D} \left| \sum_{n=0}^{M-1} e_{n+d_1} e_{n+d_2} ... e_{n+d_k} \right|,$$

where the maximum is taken over all $D = (d_1, ...d_k)$ ($d_1 < ... < d_k$ are nonnegative integers) and $M \in \mathbb{N}$ with $M + d_k \leq N$.

Well-distribution measure of E_N is:

$$W(E_N) = \max_{a,b,t} \left| \sum_{j=1}^{t} e_{a+jb} \right|,$$

where the maximum is taken over all $a \in \mathbb{Z}$, $b, t \in \mathbb{N}$ and $1 \leq a + b \leq a + tb \leq N$.

E_N is considered a **good** pseudorandom sequence,
if both $C_k(E_N)$ and $W(E_N)$ are "small" in terms of N.

This terminology is justified:
Cassaigne, Mauduit and Sárközy (2002):
for almost all $E_N = \{-1, +1\}^N$ truly random sequence
both measures are small:
$O(N^{1/2}(\log N)^c)$

Main topic of my research:
collisions and avalanche effect

Important in applications: e.g. DES

$\mathcal{S}$ is a given set
Assume that $N \in \mathbb{N}$, $\mathcal{S}$ is a given set and to each $s \in \mathcal{S}$ we assign a unique binary sequence

$$E_N = E_N(s) = (e_1, ..., e_N) \in \{-1, +1\}^N,$$

and let $\mathcal{F} = \mathcal{F}(\mathcal{S})$ denote the family of the binary sequences obtained in this way:

$$\mathcal{F} = \mathcal{F}(\mathcal{S}) = \{E_N(s) : s \in \mathcal{S}\}. \quad (1)$$

Definition 1

If $s \in \mathcal{S}, s' \in \mathcal{S}, s \neq s'$ *and*

$$E_N(s) = E_N(s'), \tag{2}$$

then (2) is said to be a **collision** *in* $\mathcal{F} = \mathcal{F}(\mathcal{S})$.
If there is no collision in $\mathcal{F} = \mathcal{F}(\mathcal{S})$, *then* $\mathcal{F}$ *is said to be* **collision free**.

In other words, $\mathcal{F} = \mathcal{F}(\mathcal{S})$ is collision free if we have $|\mathcal{F}| = |\mathcal{S}|$.

An ideally good family of pseudorandom binary sequences is collision free.

If $\mathcal{F}$ is not collision free but the number of collisions is limited $\Longrightarrow$ they do not cause many problems.

A good measure of the number of collisions is the following:

Definition 2

The **collision maximum** $M = M(\mathcal{F}, \mathcal{S})$ *is defined by*

$$M = M(\mathcal{F}, \mathcal{S}) = \max_{E_N \in \mathcal{F}} |\{s : s \in \mathcal{S}, E_N(s) = E_N\}|$$

(i.e., M *is the maximal number of elements of* S *representing the same binary sequence* E_N*).*

Definition 3

If in (1) we have $S = \{-1, +1\}^l$, *and for any* $s \in S$, *changing any element of* s *changes "many" elements of* $E_N(s)$ *(i.e., for* $s \neq s'$ *many elements of the sequences* $E_N(s)$ *and* $E_N(s')$ *are different), then we speak about* **avalanche effect**, *and we say that* $\mathcal{F} = \mathcal{F}(S)$ *possesses the* **avalanche property**.

If for any $s \in S, s' \in S, s \neq s'$ *at least* $(\frac{1}{2} - o(1))N$ *elements of* $E_N(s)$ *and* $E_N(s')$ *are different then* $\mathcal{F}$ *is said to possess* **strict avalanche property**.

To study the avalanche property, I introduced the following measure:

Definition 4

If $N \in \mathbb{N}$, $E_N = (e_1, ..., e_N) \in \{-1, 1\}^N$ *and* $E'_N = (e'_1, ..., e'_N) \in \{-1, 1\}^N$, *then the* **distance** $d(E_N, E'_N)$ *between* E_N *and* E'_N *is defined by*

$$d(E_N, E'_N) = \left|\{n : 1 \leq n \leq N, e_n \neq e'_n\}\right|.$$

Moreover, if $\mathcal{F}$ *is a family of form (1), then the* **distance minimum** $m(\mathcal{F})$ *of* $\mathcal{F}$ *is defined by*

$$m(\mathcal{F}) = \min_{\substack{s,s' \in \mathcal{S} \\ s \neq s'}} d(E_N(s), E_N(s')).$$

Applying this notion we may say that

the family $\mathcal{F}$ is collision free $\iff m(\mathcal{F}) > 0$,

and $\mathcal{F}$ possesses the strict avalanche property if

$$m(\mathcal{F}) \geq \left(\frac{1}{2} - o(1)\right) \cdot N.$$

1st construction: Legendre symbol

A good candidate for testing the measures of pseudorandomness is the **Legendre symbol**:

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & \text{if } p|a \\ +1, & \text{if } a \text{ quadratic residue mod } p \\ -1, & \text{if } a \text{ nonquadratic residue mod } p \end{cases}$$

- its random behaviour is known for long (Jacobstahl, Davenport, Bach, Peralta, Damgard, Sárközy)

Mauduit and Sárközy, 1997 :

$$e_n = \left(\frac{n}{p}\right) \quad (n = 1, 2, ..., p-1)$$

Goubin, Mauduit and Sárközy, 2004 :

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right), & \text{if } (f(n), p) = 1 \\ +1, & \text{if } p | f(n). \end{cases} \tag{3}$$

Theorem 1 (VTóth)

Let $\mathcal{S}$ be the set of polynomials $f(x) \in \mathbb{F}_p[X]$ of degree $D \geq 2$ which do not have multiple zeros. Define $E_p = E_p(f) = (e_1, ..., e_p)$ by (3) and $\mathcal{F} = \mathcal{F}(\mathcal{S})$ by (1). Then we have

$$m(\mathcal{F}) \geq \frac{1}{2}\left(p - (2D-1)p^{1/2} - 2D\right).$$

The proof is based on the theorem of Weil.

Corollary 1 (VTóth)

If $\mathcal{S}$, $\mathcal{F}$ are defined as in Theorem 1 and we also have $D < \frac{p^{1/2}}{2}$, then $\mathcal{F}$ is collision free.

Corollary 2 (VTóth)

If $\mathcal{S}$, $\mathcal{F}$ are defined as in Theorem 1 and we have $p \to +\infty$, $D = o(p^{1/2})$ then $\mathcal{F}$ possesses the strong avalanche property.

2nd construction: based on additive character

Mauduit, Rivat and Sárközy introduced the following construction in 2004:

let p be an odd prime number, $f(X) \in \mathbb{F}_p[X]$, and define $E_p = (e_1, ..., e_p)$ by

$$e_n = \begin{cases} +1, & \text{if } 0 \leq r_p(f(n)) < p/2 \\ -1, & \text{if } p/2 \leq r_p(f(n)) < p, \end{cases} \tag{4}$$

where $r_p(n)$ denotes the unique $r \in \{0, ..., p-1\}$ such that $n \equiv r \pmod p$.

- Advantages:
  - small measures
  - fast

- Disadvantages:
  - correlation measure of large order can be large (Mauduit, Rivat and Sárközy)
  - there are "many" collisions in it

"Many" collisions:
$\mathcal{S}_k = \{f(x) : f(x) \in \mathbb{F}_p[x], \deg f(x) = k\}$
$\mathcal{F}_k = \{E_p(f) = (e_1, \ldots e_p) : f \in \mathcal{S}_k\}$
If

$$\frac{k}{p(\log p)^{-1}} \to \infty,$$

then

$$M(\mathcal{F}_k, \mathcal{S}_k) \to \infty.$$

Theorem 2 (VTóth)

If p is a fixed prime and $\mathcal{F}_2$, $\mathcal{S}_2$ are defined as above then we have $M(\mathcal{F}_2, \mathcal{S}_2) \geq \lfloor \frac{1}{6} \log p \rfloor$.

It can be saved:

$$\mathcal{P}_d = \{f(x) \in \mathbb{F}_p[x] : f(x) = \sum_{i=0}^{d} a_i x^i \text{ ,ahol } a_0 = 0, a_d = 1\}$$

Theorem 3 (VTóth)

If $f(x) \in \mathcal{P}_d$, then the family of binary sequences constructed by (4) is collision free and possesses the strict avalanche property.

Java programme by Viktória Fonyó

- Goal: testing the constructions in the "real life"
  - generation of the sequences: fast
  - calculation of the measures: comparing with other constructions
  - using the sequences in Vernam cipher
- Result: they can be used easily and in a fast way in applications as well

Conclusion

- **large families** of binary sequences with strong pseudorandom properties

- mathematically **provable** nice properties

- can be used in **applications**


I. Damgard, *On the randomness Legendre and Jacobi sequences*, Lect. Notes in Comp. Sci. 403, Springer-Verlag, Berlin (1990), 163–172.


V. Fonyó, *Pszeudovéletlen sorozatok konstrukciói*, Thesis work (2017)


L. Goubin, C. Mauduit, A. Sárközy, *Construction of large families of pseudorandom binary sequences*, J. Number Theory 106 (2004), 56–69.


C. Mauduit, J. Rivat, A. Sárközy, *Construction of pseudorandom binary sequences using additive characters*, Monatshefte Math. 141 (2004), 197–208.


C. Mauduit, A. Sárközy, *On finite pseudorandom binary sequences I: The measures of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (1997) 365–377.


V. Tóth, *Collision and avalanche effect in families of pseudorandom binary sequences*, Periodica Math. Hungar. 55. (2007) 2, 185–196.

 V. Tóth, *The study of collision and avalanche effect in a family of pseudorandom binary sequences*, Periodica Math. Hungar. 59. (2009) 1, 1–8.

 V. Tóth, *Extension of the notion of collision and avalanche effect to sequences of k symbols*, Periodica Math. Hungar. 65. (2012) 2, 229–238.

 V. Tóth, *Collision and avalanche effect in pseudorandom sequences*, Annales Univ. Sci. Budapest., Sect. Comp. 41. (2013), 347–354.