Families of pseudorandom binary sequences with low cross-correlation measure

Oğuz Yayla

Johann Radon Institute for Computational and Applied Mathematics

Austrian Academy of Sciences
Linz, Austria

BalkanCryptSec 2014
İstanbul

October 16, 2014

Outline

- Introduction
- Constructed large families
- Application to cryptography

Sequences

Pseudorandom sequences

- in cryptography - a key stream of stream ciphers.
- be unpredictable and resist to known attacks.
- large linear complexity and low correlation.

Sequences

Pseudorandom sequences

- in cryptography - a key stream of stream ciphers.
- be unpredictable and resist to known attacks.
- large linear complexity and low correlation.

Family of sequences

- complex and rich structure.
- large family size, large family complexity, and low cross-correlation.

Contribution

1. Two large families of pseudorandom binary sequences with low cross-correlation measure
2. Extension of the family construction method given by K. Gyarmati, C. Mauduit and A. Sárközy.

Pseudorandomness measures

Binary sequence

$$E_N = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N.$$

- *Well distribution measure*:

$$W(E_N) = \max_{a,b,t} \left| \sum_{j=0}^{t-1} e_{a+bj} \right|,$$

where the maximum is taken over all $a \in \mathbb{N} \cup \{0\}$, $b, t \in \mathbb{N}$ such that $0 \le a \le a + b(t-1) \le N-1$

Pseudorandomness measures

- *Correlation measure of order k*

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

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ and M such that $0 \leq d_1 < d_2 < \cdots < d_k \leq N - M$.

Good pseudorandom sequences

- Cassaigne, Mauduit, and Sarkozy in 2002 proved that $W(E_N)$ and $C_k(E_N)$ are "small" for a truly random sequence $E_N \in \{-1, +1\}$,
- $W(E_N)$ and $C_k(E_N)$ (for fixed k) are $N^{1/2}$ and $N^{1/2}(\log N)^{c(k)}$
- "Good" pseudorandom sequence if both $W(E_N)$ and $C_k(E_N)$ (for small k) are small and ideally greater than $N^{1/2}$ only by at most a power of $\log N$.

Measures in cryptanalysis

- $E_N \in \{-1, +1\}$, a key stream in cryptographic applications.
- E_N must be unpredictable
- Exhaustive search on the set of all possible binary sequences $E_N \in \{-1, +1\}$ with large $W(E_N)$ (or large $C_k(E_N)$)
- Since this set is much smaller than the set of all sequences in $\{-1, +1\}^N$, the attack recovers the key if the key is not a "good" pseudorandom sequence.
- Besides a fast method of exhaustive search, one also needs a fast algorithm to generate the set of sequences with large $W(E_N)$ (or large $C_k(E_N)$).

Previous Constructions

- Mauduit and Sarkozy in 1997 showed that the Legendre symbol forms a “good” pseudorandom sequence.
- “good” pseudorandom sequences are very few
- in cryptography we generally need large families of “good” pseudorandom sequences
- Large families of “good” pseudorandom binary sequences with low well distribution and correlation measures were also constructed, see Gyarmati et al. — individual sequences
- Not enough to say that the family is good, and in many applications we need that the family has a complex and rich structure

Previous Constructions

- *Family complexity*, *collision and avalanche effect* are introduced
- Gyarmati, Mauduit and Sárközy in 2014 - *cross-correlation measure of order k* to characterize a family of sequences.
- Winterhof and Y. recently showed that the family complexity of a binary sequence can be estimated by the cross-correlation measure of its dual family

Cross-correlation measure

Definition 1

The *cross-correlation measure of order k* of a family $\mathcal{F}$ of binary sequences $E_{i,N} = (e_{i,1}, e_{i,2}, \dots, e_{i,N}) \in \{-1+1\}^N$, $i = 1, 2, \dots, |\mathcal{F}|$, is defined as

$$\Phi_k(\mathcal{F}) = \max_{M,D,I} \left| \sum_{n=1}^{M} e_{i_1,n+d_1} \cdots e_{i_k,n+d_k} \right|$$

where D denotes a k tuple $(d_1, d_2, \dots, d_k)$ of integers such that $0 \le d_1 \le d_2 \le \dots \le d_k < M + d_k \le N$ and $d_i \ne d_j$ if $E_{i,N} = E_{j,N}$ for $i \ne j$, and I denotes a k tuple $(i_1, i_2, \dots, i_k)$ in $\{1, 2, \dots, |\mathcal{F}|\}$.

Previous Result on cross-correlation

- Gyarmati, Mauduit and Sárközy show the connection between the cross-correlation measure of order k and other measures.
- Then they present two families of pseudorandom binary sequences with small cross-correlation measure.

We extend their families of pseudorandom binary sequences. We obtain larger families of pseudorandom binary sequences which have small cross correlation measure of order k.

Legendre sequence

Let $\mathcal{F}$ be the family of binary sequences $E_p(f) = (e_1, e_2, \ldots, e_p)$ assigned to the polynomial f by the formula:

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right) & \text{if } p \nmid f(n) \\ +1 & \text{if } p \mid f(n) \end{cases} \tag{1}$$

for $n = 1, 2, \ldots, p$.

Family 1

Theorem 2

Let p be a prime number and $d \in \mathbb{Z}^+$ such that $d < p^{1/2}/(20 \log p)$. Then consider all polynomials of the form

$$f(x) = (x - x_1)(x - x_2) \dots (x - x_t) \tag{2}$$

where $x_1, x_2, \dots, x_t$ are distinct elements of $\mathbb{F}_p$ and

$$x_1 + x_2 + \dots + x_t = 0 \tag{3}$$

such that $1 \le t \le d$. Let $\mathcal{F}_1$ be the family of binary sequences $E_p(f) = (e_1, e_2, \dots, e_p)$ assigned to the polynomial f by the formula (1).

Family 1

(i) $\phi_2(\mathcal{F}_1) < 20dp^{1/2}\log p$. If the second order cross-correlation measure only, then

$$|\mathcal{F}_1| = \sum_{t=1}^{d} \frac{1}{t}\binom{p-1}{t-1}. \tag{4}$$

(ii) If k and t are odd integers for all $f \in \mathcal{F}_1$, then $\phi_k(\mathcal{F}_1) < 10kdp^{1/2}\log p$. In this case,

$$|\mathcal{F}_1| = \sum_{\substack{t=1 \\ t\text{-}odd}}^{d} \frac{1}{t}\binom{p-1}{t-1}.$$

(iii) If k is an odd integer and $t = 2$ for all $f \in \mathcal{F}$, then $\phi_k(\mathcal{F}_1) < 20kp^{1/2}\log p$. And, $|\mathcal{F}_1| = \dfrac{p-1}{2}$.

A corollary

Corollary 3

Consider the polynomials of the form

$$f(x) = (x - x_1)^{s_1}(x - x_2)^{s_2} \cdots (x - x_t)^{s_t}$$

where $x_1, x_2, \ldots, x_t$ are distinct elements of $\mathbb{F}_p$ such that $1 \leq s_1 + s_2 + \ldots + s_t \leq d$.
Let k and $\deg(f)$ be odd integers for $f \in \mathcal{F}$. Then we have $\phi_k(\mathcal{F}) < 10kdp^{1/2} \log p$.

Much larger, but collision

Larger:

$$h(x) = (x - x_1)^2 (x - x_2) \in \mathbb{F}_p[x]$$

But collisions:

$f(x) = (x - x_1)^3 (x - x_2)^5 (x - x_3)$ and $g(x) = (x - x_1)(x - x_2)(x - x_3)^7$

Family 2

Theorem 4

Consider all irreducible polynomials $f(x) \in F_p[x]$ of the form

$$f(x) = x^t + a_2 x^{t-2} + a_3 x^{t-3} + \cdots + a_t \tag{5}$$

for some integer $2 \leq t \leq d$ and let $\mathcal{F}_2$ family of the binary sequences generated (1). Then,

$$\phi_k(\mathcal{F}_2) < 9kdp^{1/2} \log p \tag{6}$$

for all $k = 2, 3, \ldots, p-1$. $|\mathcal{F}_2| \geq \sum_{t=2}^{d} p^{[t/3]-1}$.

How to apply?

- In cryptographic applications we need large key space, ie large family size.
- Family sizes of the constructed families are exponentially grow by the degree d of the seed polynomial f
- To guarantee the good pseudorandom properties of the constructed sequences we choose the degree from the interval $3 \leq d \leq p^{1/4}$.
- Choose d near to the lower end so that the sequences possess better pseudorandom properties.

Shortening

- Shortening the sequence at a position $M < p$, may cause the sequence to loose the pseudorandom properties.
- if $M \geq \lceil p^{\frac{1}{4\sqrt{e}}} \rceil$ it is known that the sequence still preserves its pseudorandom properties

An example

- choose $p = 10^{10} + 19$
- the family size of $\mathcal{F}_1$ becomes at least 2^{125}, 2^{247}, and 2^{541} for $d = 5, 9$, and 19
- the family size of $\mathcal{F}_2$ becomes at least 2^{132}, 2^{265}, and 2^{532} for $d = 15, 27$, and 52
- shortened at a position $M \geq \lceil p^{\frac{1}{4\sqrt{e}}} \rceil \approx 100$, it still has the good properties.
- use such sequences without shortening e.g. for an encryption of a video steam having block length p (≈ 1 gigabyte).

...

Thanks.