On the maximum order complexity of subsequences of the Thue-Morse and Rudin-Shapiro sequence along squares

Arne Winterhof

Austrian Academy of Sciences
Johann Radon Institute for Computational and Applied Mathematics
Linz

UDT, Luminy, Oct. 2, 2018

How do we measure the suitability of a binary sequence for cryptography?

How do we measure the suitability of a binary sequence for cryptography?

How do we detect 'bad' cryptographic sequences?

How do we find 'good' cryptographic sequences?

How do we measure the suitability of a binary sequence for cryptography?

Study several quality measures:

- **linear complexity**
- **maximum order complexity**
- **correlation measure**
- Gowers norm
- expansion complexity
- ...

Studied sequences

- Thue-Morse sequence (Rudin-Shapiro sequence)
- pattern sequences
- (non-periodic) automatic sequences
- subsequence of Thue-Morse along the squares
- Legendre sequence (with polynomials)
- ...

Linear Complexity

For a positive integer N the Nth linear complexity $L(s_n, N)$ of a sequence (s_n) over $\mathbb{F}_2$ is the smallest positive integer L such that there are constants $c_0, \ldots, c_{L-1} \in \mathbb{F}$ satisfying

$$s_{n+L} = c_{L-1}s_{n+L-1} + \ldots + c_0 s_n,$$

$0 \leq n \leq N - L - 1.$

Linear Complexity

For a positive integer N the Nth linear complexity $L(s_n, N)$ of a sequence (s_n) over $\mathbb{F}_2$ is the smallest positive integer L such that there are constants $c_0, \ldots, c_{L-1} \in \mathbb{F}$ satisfying

$$s_{n+L} = c_{L-1}s_{n+L-1} + \ldots + c_0 s_n,$$

$0 \leq n \leq N - L - 1$.

- small linear complexity implies predictability and thus unsuitability in cryptography
- there are predictable sequences with large linear complexity
- we need further measures for the unpredictability of a sequence

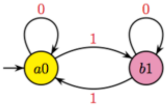
Automatic sequences:

- high linear complexity
- still predictable

2-automatic sequences

Thue-Morse sequence $s_0 = 0,\ s_{2n+1} = s_n + 1,\ s_{2n} = s_n,\ n = 0, 1, \dots$

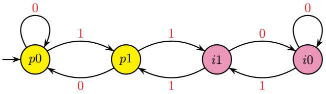
Input $n = (n_0 n_1 \dots n_j)_2$



Output:0, 1

$011010011001011 0\dots$ $n = 5 = (101)_2$

Rudin-Shapiro



$0001001000011101\dots$

Christol's theorem

generating function $G(x)$ of (s_n):

$$G(x) = \sum_{n=0}^{\infty} s_n x^n$$

(s_n) is 2-automatic over $\mathbb{F}_2$ if and only if $G(x)$ is algebraic over $\mathbb{F}_2[x]$.

Example. Thue-Morse sequence

$$(x+1)^3 G(x)^2 + (x+1)^2 G(x) + x = 0$$

$G(x)$ rational if and only if (s_n) is ultimately periodic.

Linear complexity of Thue-Morse sequence

Mérai/W., 2018:

$$t_{n+L} = \sum_{\ell=0}^{L-1} c_\ell t_{n+\ell} = 0, \quad 0 \le n \le N-1-L$$

$$f(x) = \sum_{n=0}^{L} c_\ell x^{L-n}, \quad c_L = 1$$

$$g(x) := f(x)G(x) \bmod x^N$$

$\deg(g) < L$, $\deg(f) \le L$

$$(x+1)^3 g(x)^2 + (x+1)^2 f(x) g(x) + x f(x)^2 = k(x) x^N, \quad k(x) \neq 0$$

$$L(t_n, N) \ge \frac{N-1}{2}$$

- A different method gives $L(t_n, N) = 2\lfloor (N+2)/4 \rfloor$.
- The method works for any (non-periodic) automatic sequence (if we know $h \neq 0$ with $h(x, G(x)) = 0$).
- What measure(s) can be used to see that the Thue-Morse sequence (other automatic sequences) is (are) predictable?
- What can we say about (non-automatic) subsequences of automatic sequences?
- Are there any sequences with good results under all commonly used measures of pseudorandomness?

Correlation measure of order k

Mauduit/Sárközy, 1997:
The correlation measure of order k of a binary sequence $(s_n)_{n=0}^{T-1}$ of length T is introduced as

$$C_k(s_n) = \max_{M,D} \left| \sum_{n=1}^{M} (-1)^{s_{n+d_1}} \cdots (-1)^{s_{n+d_k}} \right|, \quad k \geq 1,$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ with non-negative integers $d_1 < d_2 < \cdots < d_k$ and M such that $M - 1 + d_k \leq T - 1$.

Correlation measure of order k

Mauduit/Sárközy, 1997:
The correlation measure of order k of a binary sequence $(s_n)_{n=0}^{T-1}$ of length T is introduced as

$$C_k(s_n) = \max_{M,D} \left| \sum_{n=1}^{M} (-1)^{s_{n+d_1}} \cdots (-1)^{s_{n+d_k}} \right|, \quad k \geq 1,$$

where the maximum is taken over all $D = (d_1, d_2, \ldots, d_k)$ with non-negative integers $d_1 < d_2 < \cdots < d_k$ and M such that $M - 1 + d_k \leq T - 1$.

Mauduit/Sárközy, 1998: $C_2(t_n) \geq N/12$, $N \geq 5$
Mérai/W. 2018: lower bounds on $C_2(s_n)$ for any automatic sequence.

Relation between linear complexity and correlation measure

Brandstätter/W., 2006:

$$L(s_n, N) \geq N - \max_{1 \leq k \leq L(s_n,N)+1} C_k(s_n), \quad 2 \leq N \leq T - 1.$$

Example: Legendre sequence

Let $p > 2$ be a prime. The *Legendre-sequence* (ℓ_n) is defined by

$$\ell_n = \begin{cases} 1, & \left(\frac{n}{p}\right) = -1, \\ 0, & \text{otherwise}, \end{cases} \qquad n \geq 0,$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre-symbol.

Example: Legendre sequence

Let $p > 2$ be a prime. The *Legendre-sequence* (ℓ_n) is defined by

$$\ell_n = \begin{cases} 1, & \left(\frac{n}{p}\right) = -1, \\ 0, & \text{otherwise}, \end{cases} \qquad n \geq 0,$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre-symbol.

The Nth linear complexity of the Legendre sequence satisfies

$$L(\ell_n, N) \gg \frac{\min\{N, p\}}{p^{1/2} \log p}, \quad N \geq 1.$$

$$L(s_n,N)\geq N-\max_{1\leq k\leq L(s_n,N)+1} C_k(s_n),\quad 2\leq N\leq T-1.$$

Mauduit/Sárközy, 1997:

$$C_k(\ell_n)\ll kp^{1/2}\log p$$

$$L(\ell_n,N)\gg \frac{\min\{N,p\}}{p^{1/2}\log p},\quad N\geq 1.$$

Maximum order complexity

Maximum order complexity: Smallest $M = M(s_n, N)$ with

$$s_{n+M} = f(s_{n+M-1}, \ldots, s_n), \quad 0 \leq n \leq N - M - 1.$$

- finer than linear complexity
- much more complicated to calculate

Maximum order complexity vs. correlation measure

Isik, W., 2017: For any binary sequence (s_n) we have

$$M(s_n,N) \geq N - 2^{M(s_n,N)+1} \max_{1\leq k\leq M(s_n,N)+1} C_k(s_n,N), \quad N \geq 1.$$

Maximum order complexity vs. correlation measure

Isik, W., 2017: For any binary sequence (s_n) we have

$$M(s_n, N) \geq N - 2^{M(s_n,N)+1} \max_{1 \leq k \leq M(s_n,N)+1} C_k(s_n, N), \quad N \geq 1.$$

Example. Legendre sequence

$$C_k(\ell_n) \ll kp^{1/2} \log p$$

$$M(\ell_n, N) \geq \log(\min\{N, p\}/p^{1/2}) + O(\log \log p)$$

typical: $M(s_n, N) \approx \log N$, $C_k(s_n) \approx N^{1/2} \log^{c(k)} N$

Maximum-order complexity of Thue-Morse sequence

Sun/W., submitted: $M(t_n) \geq N/5$

- Thue-Morse sequence is predictable since $C_2(t_n)$ is large.
- still $M(t_n)$ is very large

Maximum-order complexity of Thue-Morse sequence

Sun/W., submitted: $M(t_n) \geq N/5$

- Thue-Morse sequence is predictable since $C_2(t_n)$ is large.
- still $M(t_n)$ is very large

- Do certain subsequences of the Thue-Morse sequence inherit a large maximum order complexity?
- For example, the Thue-Morse sequence along the squares is not automatic and normal (Drmota/Mauduit/Rivat).

Thue-Morse sequence along squares

$s_n = t_{n^2},\ n = 0, 1, \ldots$

Sun/W., submitted: $M(s_n) \gg N^{1/2}$

Idea of proof: $5 \cdot 2^\ell < N \leq 5 \cdot 2^{\ell+1}$
Verify

$$t_{(i+2^{\ell+1})^2} = t_{(i+2^{\ell+2})^2}, \quad i = 0, 1, \ldots, \left\lfloor \sqrt{2^{\ell+2} - 1} \right\rfloor$$

and

$$t_{(2^\ell+2^{\ell+1})^2} \neq t_{(2^\ell+2^{\ell+2})^2}.$$

Assume $M \leq \left\lfloor \sqrt{2^{\ell+2} - 1} \right\rfloor + 1$ such that

$$t_{(i+M)^2} = f(t_{i^2}, \ldots, t_{(i+M-1)^2}), \quad i = 0, 1, \ldots, N - M - 1$$

we get a contradiction.

A generalization

The pattern sequences $\mathcal{P}_k = (p_n)_{n=0}^{\infty}$ is defined by

$$p_n \equiv s_k(n) \bmod 2,$$

where $P_k = 11\ldots1 \in \mathbb{F}_2^k$ is the all 1 pattern of length k and $s_k(n)$ is the number of occurrences of P_k in the binary representation of n.

$k = 1$: Thue-Morse sequence
$k = 2$: Rudin-Shapiro sequence

$$M(\mathcal{P}_k', N) \gg N^{1/2}, \quad N \geq 2^{2k+2}.$$

Thank you for your attention!