Vladimir S. Anashin

p-Adic Dynamical Systems and Cryptography

T-function=triangle Boolean mapping=determined function:

$$(\alpha_0^{\downarrow}, \alpha_1^{\downarrow}, \alpha_2^{\downarrow}, \ldots) \mapsto (\Phi_0(\alpha_0^{\downarrow}), \Phi_1(\alpha_0^{\downarrow}, \alpha_1^{\downarrow}), \Phi_2(\alpha_0^{\downarrow}, \alpha_1^{\downarrow}, \alpha_2^{\downarrow}), \ldots),$$

where $\alpha_i \in \mathbb{B}^m$, Boolean columnar m-dimensional vector;
$\Phi_i \colon (\mathbb{B}^m)^{(i+1)} \to \mathbb{B}^n$ maps $(i+1)$ Boolean columnar m-dimensional vectors to n-dimensional Boolean vector; $\mathbb{B} = \{0, 1\}$.

Example: A Stream cipher. $\alpha_j, \gamma_j, \zeta_j \in \{0, 1\}$.

Plain text:	α_0	α_1	α_2	$\ldots$
Addition mod2:	$\bigoplus$			
Key stream:	γ_0	γ_1	γ_2	$\ldots$
Encrypted text:	$\zeta_0 = \alpha_0 \oplus \gamma_0$	$\zeta_1 = \alpha_1 \oplus \gamma_1$	$\zeta_2 = \alpha_2 \oplus \gamma_2$	$\ldots$

Less trivial example: Plain-text-dependent cipher

Plain text:	α_0	α_1	α_2	$\dots$
Cipher	γ_0^K	$\gamma_1^K(\cdot)$	$\gamma_2^K(\cdot,\cdot)$	$\dots$
Encrypted text:	$\alpha_0 \oplus \gamma_0^K$	$\alpha_1 \oplus \gamma_1^K(\alpha_0)$	$\alpha_2 \oplus \gamma_2^K(\alpha_0,\alpha_1)$	$\dots$

Yet another example: Integer addition

	0	1	1	1
+				
	0	0	1	1
	1	0	1	0

Dynamical system: $\langle \mathfrak{X}, \mu, \mathbf{f} \rangle$, where $\mathfrak{X}$ is a phase space (usually a metric space), μ is a measure on $\mathfrak{X}$ (e.g., probabilistic); $\mathbf{f}\colon \mathfrak{X} \to \mathfrak{X}$ is a measurable mapping (usually, continuos). A trajectory:

$$\mathbf{x}_0, \mathbf{x}_1 = \mathbf{f}(\mathbf{x}_0), \dots, \mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i), \dots .$$

An example: Bernoulli shift (=doubling map) $\mathfrak{X} = [0, 1]$ is the real unit interval; μ is the Lebesgue measure; $\mathbf{f}(\mathbf{x}) = 2 \cdot \mathbf{x} \pmod 1$ is fractional part of $2 \cdot \mathbf{x}$. This is a chaotic system!

Yet another example: Logistic map $\mathbf{f}(\mathbf{x}) = 4 \cdot \mathbf{x} \cdot (1 - \mathbf{x}) \pmod 1$

What in common?

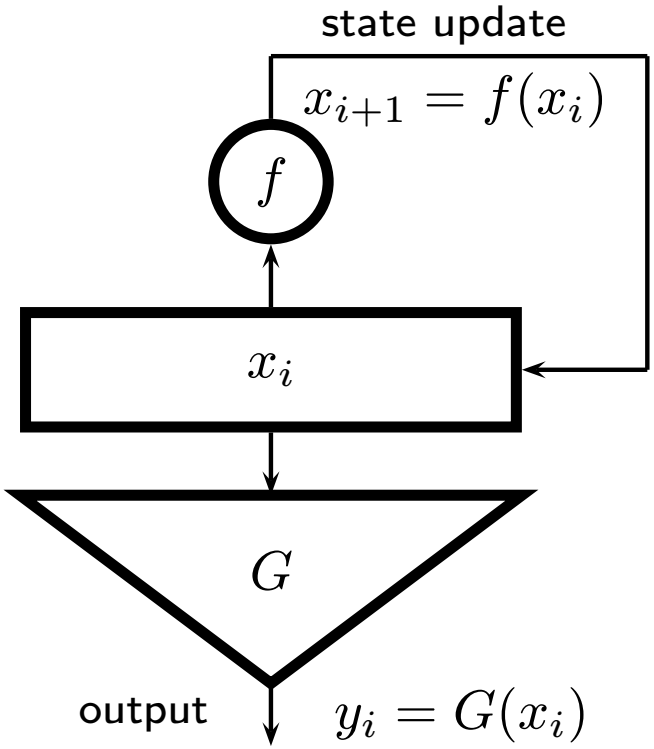
Stream cipher. $\alpha_j, \gamma_j, \zeta_j \in \{0, 1\}$.

Plain text:	α_0	α_1	α_2	$\dots$
Addition mod2:	$\bigoplus$			
Key stream:	γ_0	γ_1	γ_2	$\dots$
Encrypted text:	$\zeta_0 = \alpha_0 \oplus \gamma_0$	$\zeta_1 = \alpha_1 \oplus \gamma_1$	$\zeta_2 = \alpha_2 \oplus \gamma_2$	$\dots$

Shannon's Theorem $\Rightarrow$ *the cipher is secure whenever one chooses key stream at random;* i.e., the key stream must be a sequence of i.i.d. random variables valuated in $\{0, 1\}$.

Given a family $\mathcal{T}$ of statistical tests, a *pseudorandom* sequence (with respect to $\mathcal{T}$) is the one that passes all the tests of $\mathcal{T}$. Assuming an *adversary* can use only the tests of $\mathcal{T}$, he can not distinguish a pseudorandom sequence from a truly random one.

Pseudorandom number generator (PRNG): A finite automaton with internal alphabet A and output alphabet B. Here $f\colon A \to A$ is the state update function, $G\colon A \to B$ is the output function; usually a *key* is the initial state (*a seed*) x_0. A key is the only information that is not known to an adversary.



A secure PRNG must meet the following conditions:

- For (almost) all keys the output sequences must be pseudorandom (i.e., undistinguishable from a truly random)
- Given a segment $y_j, y_{j+1}, \dots, y_{j+s-1}$ of the output, finding the corresponding key x_0 must be infeasible.

Note. Usually length s of the output is assumed to be restricted by a polynomial in $\log |N|$.

Example. BBS generator: $f(x) = x^2 \bmod M$, $M = P \cdot Q$, the primes P, Q are not known to an adversary, $G(x) = \delta_0(x)$ (the least significant bit of x). *Conjecturing no algorithm could factorize* M in polynomial in $\log M$ time, the generator is secure; no output of polynomial in $\log M$ length could be distinguished from random in polynomial in $\log M$ time.

Note: The conjecture does not hold for quantum algorithms.

Drawback: BBS generator is too slow for practical purposes.

In quest for a fast secure PRNG. Dynamical systems theory prompts a very natural approach: Let $\langle \mathfrak{X}, \mu, \mathbf{f} \rangle$ be a dynamical system with discrete time. Take a point $\mathbf{x}_0 \in \mathfrak{X}$ as a key, use as a source of pseudorandomness the trajectory $\mathbf{x}_0, \mathbf{x}_1 = \mathbf{f}(\mathbf{x}_0), \ldots, \mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i), \ldots$. Questions:

1. How to implement this on a digital computer?
2. What will be the performance?
3. How pseudorandom is the so produced sequence?
4. Is the corresponding generator secure?

Chaos-based cryptography is based on a very natural mood — take a chaotic map $\mathbf{f}$ and discretize it! Note: the result of an ad hoc approach could be quite unexpected.

Example. A discretized version of the doubling map (Bernoulli shift) $\mathbf{f}(\mathbf{x}) = (2 \cdot \mathbf{x}) \bmod 1$ is $x_{i+1} \equiv 2 \cdot x_i \pmod{2^n}$. Becomes 0 after at most n iterations!!!

"Despite a huge number of papers published in the field of chaos-based cryptography, the impact that this research has made on conventional cryptography is rather marginal. This is due to two reasons:

- First, almost all chaos-based cryptographic algorithms use dynamical systems defined on the set of real numbers, and therefore are difficult for practical realization and circuit implementation.
- Second, security and performance of almost all proposed chaos-based methods are not analyzed in terms of the techniques developed in cryptography. Moreover, most of the proposed methods generate cryptographically weak and slow algorithms."*

*L. Kocarev. 'Chaos-Based Cryptography: A Brief Overview', in: Circuits and Systems IEEE Magazine. Vol.1, No. 3, 2001

Aiming at practical realization. Both f and G *must be* compositions of basic microprocessor instructions (operations), which include:

- integer arithmetic operations (addition, multiplication,...)
- bitwise logical operations (OR, XOR, AND, NOT)
- machine operations (shifts, masking, cyclic shifts).

Note. Let $z = \delta_0(z) + \delta_1(z) \cdot 2 + \delta_2(z) \cdot 2^2 + \delta_3(z) \cdot 2^3 + \cdots$ be a base-2 expansion for $z \in \mathbb{N}_0$. Then

- $\delta_j(y \,\mathsf{XOR}\, z) \equiv \delta_j(y) + \delta_j(z) \pmod 2$, bitwise addition modulo 2
- $y \,\mathsf{AND}\, z$ is a bitwise multiplication modulo 2
- $\lfloor \frac{z}{2} \rfloor$ is a shift towards less significant bits
- $2 \cdot z$ is a shift towards more significant bits
- $y \,\mathsf{AND}\, z$ is masking of z with the mask y; in particular, reduction modulo 2^k is just $z \pmod{2^k} = z \,\mathsf{AND}(2^k - 1)$

Observation 1. All these operations, with the only exception of cyclic shifts, are defined on the space $\mathbb{Z}_2$ of all 2-adic integers.

The space $\mathbb{Z}_2$ could be thought of as a set of all countable infinite binary sequences.

Addition:

$$\begin{array}{rcccc} & \ldots 1 & 1 & 1 & 1 \\ + & & & & \\ & \ldots 0 & 0 & 0 & 1 \\ \hline & \ldots 0 & 0 & 0 & 0 \end{array}$$

Hence: $\ldots 11111 = -1$.

Multiplication:

$$
\begin{array}{rcccccc}
 & \ldots 0 & 1 & 0 & 1 & 0 & 1 \\
\times & & & & & & \\
 & \ldots 0 & 0 & 0 & 0 & 1 & 1 \\
\hline
 & \ldots 0 & 1 & 0 & 1 & 0 & 1 \\
+ & & & & & & \\
 & \ldots 1 & 0 & 1 & 0 & 1 & \\
\hline
 & \ldots 1 & 1 & 1 & 1 & 1 & 1
\end{array}
$$

Hence, $\ldots 1010101 = -\frac{1}{3}$.

Sequences with only finite number of 1's correspond to non-negative rational integers in their base-2 expansions, sequences with only finite number of 0's correspond to negative rational integers, while eventually periodic sequences correspond to rational numbers represented by irreducible fractions with odd denominators:

$$\begin{aligned} \ldots 00011 &= 3, \\ \ldots 111111 &= -1, \\ \ldots 111100 &= -4, \\ \ldots 1010101 &= -\tfrac{1}{3} \end{aligned}$$

Distance: $d_2(-1,3) = \|(-1)-3\|_2 = \|-4\|_2 = \frac{1}{2^2} = \frac{1}{4}$.

That is: $-1 \equiv 3 \pmod 4$; $-1 \not\equiv 3 \pmod 8$.

Example:

$$1, 3, 7, 15, 31, \ldots, 2^n - 1, \ldots \underset{d_2}{\longrightarrow} -1,$$

Observation 2. These operations (with the exception of cyclic shifts) could be uniquelly expanded to continuous $\mathbb{Z}_2$-valued functions of 2-adic integer arguments.

Observation 3. All these functions (with the exception of those defined by shifts towards less significant bits) are T-functions.
Example: $2 = 1 \text{ XOR } 3 = 2 \text{ AND } 7 = \text{ NOT } 13 \pmod 8$,

Observation 4. All these functions (with the exception of those defined by shifts towards less significant bits) satisfy Lipschitz condition with coefficient 1 with respect to a 2-adic metric.
Note. "$F(a) \equiv F(b) \pmod{2^k}$ whenever $a \equiv b \pmod{2^k}$" $\Leftrightarrow$ "$\|F(a) - F(b)\|_2 \le \|a - b\|_2$". By this reason, we call F *compatible.*

- A computer works with approximations of 2-adic integers up to a certain precision with respect to a 2-adic metric.
- One may use also subtraction, division raising to a power: $3^{-1} \equiv 11 \equiv -5 \pmod{16}$, $3^{-\frac{1}{3}} \equiv 3^{11} \equiv 3^{-5} \equiv 11 \pmod{16}$

From a computer's view, the following function is well defined:

$$g(x) = \left(1 - 2 \cdot \frac{x \text{ AND } x^2 + x^3 \text{ OR } x^4}{3 - 4 \cdot (5 + 6x^5)^{x^6 \text{ XOR } x^7}}\right)^{7 - \frac{8x^8}{9+10x^9}}$$

A computer evaluates this function correctly within any possible 2-adic precision he can achieve.

Natural metric of computer's world is 2-adic, non-Archimedean!

Satisfying cryptographic demands. A discretization of a dynamical system is defined on finite set N; whence, all orbits are (eventually) periodic. For many chaos-based cryptosystems a point too often falls into unexpectedly short periods, thus making a cipher insecure. A period must be long! Let make it the longest, $|N|$

Definition 1. A compatible mapping $F\colon \mathbb{Z}_p \to \mathbb{Z}_p$ is called *transitive modulo* p^k iff the induced mapping $x \mapsto F(x) \pmod{p^k}$ is a single cycle permutation on $\mathbb{Z}/p^k$.

Note. From this very moment we are mainly focused on the case $|N| = 2^k$. However, further results could be expanded to the case of a state set of arbitrary order; in particular for a p-adic case, p odd.

Theorem 1. *A compatible mapping $F\colon \mathbb{Z}_p \to \mathbb{Z}_p$ is transitive modulo p^k for all $k = 1, 2, 3, \ldots$ iff it is ergodic with respect to the Haar measure μ on $\mathbb{Z}_p$* (we normalize μ so that $\mu(\mathbb{Z}_p) = 1$).

How to determine ergodic functions among all compatible ones?

Any function $F\colon \mathbb{Z}_p \to \mathbb{Z}_p$ could be represented by Mahler's interpolation series: $F(x) = \sum_{j=0}^{\infty} c_j \binom{x}{j}$ for suitable $c_j \in \mathbb{Z}_p$. Recall

$$\binom{x}{i} = \begin{cases} \dfrac{x(x-1)\cdots(x-i+1)}{i!}, & \text{for } i = 1, 2, \ldots; \\ 1, & \text{for } i = 0. \end{cases}$$

An attempt to find an answer in terms of Mahler's interpolation series looks quite natural!

Theorem 2. *For $p = 2$ the function $F\colon \mathbb{Z}_p \to \mathbb{Z}_p$ is compatible and ergodic* $\Leftrightarrow$

$$F(x) = 1 + x + \sum_{i=1}^{\infty} c_i \cdot p^{\lfloor \log_p(i+1) \rfloor + 1} \binom{x}{i},$$

for suitable $c_i \in \mathbb{Z}_p$. (Note: For $p \neq 2$ one has $\Leftarrow$, and <u>not</u> $\Leftrightarrow$).

This theorem works well for 'analytic-looking' functions and those which could be (explicitely) expressed as Mahler's series:

Examples. For $p = 2$ the following is true:

1. The function $F(x) = a \cdot x + a^x$ is ergodic $\Leftrightarrow$ a is odd
2. The function $F(x) = -\frac{1}{2x+1} - x$ is ergodic
3. (M.V. Larin) A polynomial with integer coefficients is ergodic $\Leftrightarrow$ it is transitive modulo 8.
4. The function $F(x) = a_0 + b_1 \cdot (x \,\mathsf{XOR}\, a_1) + b_2 \cdot (x \,\mathsf{XOR}\, a_2) + \cdots$ is ergodic $\Leftrightarrow$ it is transitive modulo 4.

In fact, examples 1,2, and 3 belong to a special interesting class $\mathcal{B}$ of functions. Let

$$x^{\underline{0}} = 1,\ x^{\underline{1}} = x,\ x^{\underline{2}} = x(x-1), \ldots,\ x^{\underline{i}} = x(x-1)\cdots(x-i+1), \ldots,$$

be descending factorial powers. Let $\mathcal{B}$ be a class of all functions $F\colon \mathbb{Z}_p \to \mathbb{Z}_p$ that could be represented as $F(x) = \sum_{i=0}^{\infty} c_i \cdot x^{\underline{i}}$ for suitable $c_0, c_1, \cdots \in \mathbb{Z}_p$. The class of all functions $\mathbb{Z}_p \to \mathbb{Z}_p$ is endowed with a natural (non-Archimedean) metric
$D(U,V) = \max\{\|U(z) - V(z)\|_p \colon z \in \mathbb{Z}_p\}$

Theorem 3. *$\mathcal{B}$ is a ring, and a separable and complete (with respect to D) metric space of functions that are compatible and uniformly differentiable everywhere on $\mathbb{Z}_2$. The class $\mathcal{B}$ is closed with respect to compositions and with respect to derivations. The set $\mathcal{P}$ of all polynomials over $\mathbb{Z}$ is a dence subset of $\mathcal{B}$. All analytic on $\mathbb{Z}_p$ functions are a proper subclass of $\mathcal{B}$. A function $F \in \mathcal{B}$ is ergodic iff F is transitive modulo p^2 $(p > 3)$, or modulo p^3 $(p \le 3)$*

The nature of example 4 differs from the one of examples 1,2, and 3. To understand this nature we need some notion. For $\mathbf{a}=(a_1,\dots,a_n),\mathbf{b}=(b_1,\dots,b_n)\in\mathbb{Q}_p^n$ we write $\mathbf{a}\equiv\mathbf{b}\pmod{p^s}$ iff $\|a_i-b_i\|_p\le p^{-s}$ and define $\|\mathbf{a}\|_p=\max\{\|a_i\|_p\colon i=0,1,2,\dots\}$.

Definition 2. $F=(f_1,\dots,f_m)\colon\mathbb{Z}_p^n\to\mathbb{Z}_p^m$ is *differentiable modulo* p^k at the point $\mathbf{u}=(u_1,\dots,u_n)\in\mathbb{Z}_p^n$ iff $\exists N\in\mathbb{N}$ and an $n\times m$ matrix $F'_k(\mathbf{u})$ over $\mathbb{Q}_p$ (*the Jacobi matrix modulo* p^k) such that $\forall K\ge N$ and $\forall\mathbf{h}=(h_1,\dots,h_n)\in\mathbb{Z}_p^n$, $\|\mathbf{h}\|_p=p^{-K}$ holds

$$F(\mathbf{u}+\mathbf{h})\equiv F(\mathbf{u})+\mathbf{h}\cdot F'_k(\mathbf{u})\pmod{p^{k+K}}. \tag{1}$$

Uniform differentiability modulo p^k *on* $\mathbb{Z}_p^n$: $N_k(F)=\min K$ such that (1) holds simultaneously for all $\mathbf{u},\mathbf{h}\in\mathbb{Z}_p^n$, $\|\mathbf{h}\|_p\le p^{-K}$.

Compare: F is differentiable $\Leftrightarrow$

$$F(\mathbf{u}+\mathbf{h})\equiv F(\mathbf{u})+\mathbf{h}\cdot F'_k(\mathbf{u})\pmod{p^{\Psi(K)}},$$

where $\Psi(K)$ tends to ∞ faster than K.

Univariate case:

$$\frac{F(\mathbf{u}+\mathbf{h})-F(\mathbf{u})}{\mathbf{h}} \approx F'_k(\mathbf{u})$$

$\approx$ with arbitrarily high precision $\Rightarrow$ differentiability

$\approx$ with precision not worse than $p^{-k} \Rightarrow$ differentiability mod p^k

Rules of derivation modulo p^k: replace "$=$" with "$\equiv$" in formulae.

Examples.

- $F(x,y) = x \,\mathsf{XOR}\, y$ is not uniformly differentiable on $\mathbb{Z}_2^2$; yet it is uniformly differentiable modulo 2,

$$F'_1(x,y) \equiv \begin{pmatrix} 1 \\ 1 \end{pmatrix} \pmod 2, \text{ i.e., } d_1F(x,y) \equiv d_1x + d_1y \pmod 2$$

- $F(x) = x \bmod p^n$ is uniformly differentiable on $\mathbb{Z}_p$, $F'(x) = 0$.
- $F(x,y) = x \,\mathsf{OR}\, y$ is differentiable modulo 2 at no point of $\mathbb{Z}_2^{(2)}$; it is uniformly differentiable with respect to x for each $y \in \mathbb{Z}$; its derivative is 1 for $y \geq 0$, and it is 0 in the opposite case.

Theorem 4. *Let a compatible function $F\colon \mathbb{Z}_p \to \mathbb{Z}_p$ be uniformly differentiable modulo p^2. Then F is ergodic if and only if it is transitive modulo $p^{N_2(F)+1}$ for odd prime p or, respectively, modulo $2^{N_2(F)+2}$ for $p = 2$.* (Recall Definition 2.)

Example. (Klimov-Shamir *) The function $x + (x^2 \,\mathsf{OR}\, 5)$ is ergodic.

Proof. The function $F(x) = x + (x^2 \,\mathsf{OR}\, 5)$ is uniformly differentiable on $\mathbb{Z}_2$; thus, it is uniformly differentiable modulo 4:
$F'(x) = 1 + 2x \cdot (x \,\mathsf{OR}\, 5)' = 1 + 2x$, and $N_2(f) = 3$. Now to prove that f is ergodic, in view of Theorem 4 it sufficies to demonstrate that f induces a permutation with a single cycle on $\mathbb{Z}/32$. One verifies this by direct calculations. □

*'A new class of invertible mappings', in: *Cryptographic Hardware and Embedded Systems 2002* (B.S.Kaliski Jr.et al., eds.)), Lect. Notes in Comp. Sci.,Vol. 2523, Springer-Verlag, 2003, pp.470–483

Theorem 5. *A polynomial $F(x) \in \mathbb{Q}_p[x]$ is integer-valued* (i.e., *$F(\mathbb{Z}_p) \subset \mathbb{Z}_p$), compatible and ergodic iff the mapping*

$$z \mapsto F(z) \bmod p^{\lfloor \log_p(\deg F)\rfloor+3},$$

with z ranging over $\{0, 1, \dots, p^{\lfloor \log_p(\deg F)\rfloor+3} - 1\}$, defines a compatible and transitive function on the residue class ring $\mathbb{Z}/p^{\lfloor \log_p(\deg F)\rfloor+3}$. Loosely speaking, to determine whether a polynomial $F(x)$ with rational (and not necessarily integer) coefficients is ergodic, one has to make $\approx p^3 \cdot \deg F$ evaluations.

Theorem 6. *Denote $\Delta U(x) = U(x+1) - U(x)$. For $p = 2$ the function $F\colon \mathbb{Z}_p \to \mathbb{Z}_p$ is compatible and ergodic $\Leftrightarrow$ $F(x) = 1 + x + p \cdot \Delta U(x)$ for a compatible function $U\colon \mathbb{Z}_p \to \mathbb{Z}_p$. For $p \neq 2$ only $\Leftarrow$ is true.*

Theorem 7. *Let the function $F = (f_1, \dots, f_n)\colon \mathbb{Z}_p^n \to \mathbb{Z}_p^n$ be compatible, ergodic, and uniformly differentiable modulo p on $\mathbb{Z}_p$. Then $n = 1$.* (Non-differentiable mod$p$ ones do exist for $n > 1$)

What output functions do we need, if any? Our PRNG generates a sequence $\{y_i\}$ according to the recurrence law

$$\begin{cases} x_{i+1} &= f(x_i) \equiv F(x_i) \pmod{2^n}; \\ y_i &= G(x_i), \end{cases}$$

where F is compatible and ergodic on $\mathbb{Z}_2$. Thus, the sequence $\{x_i \in \mathbb{Z}/2^n\}$ is *strictly uniformly distributed*; i.e., it is periodic; the length of its period is the longest possible, 2^n; each element of $\mathbb{Z}/2^n$ occurs at the period exactly once. However, given $z \in \mathbb{Z}/2^n$, it is NOT computationally difficult to find a unique $x \in \mathbb{Z}/2^n$ to satisfy $z = F(x) \pmod{2^n}$. Hence, in case G is, say, the identity map, the system is insecure! Thus, G must simultaneously:

1) make it secure: given an output y_i, make it computationally difficult to find x_i that satisfy $y_i = G(x_i)$, and

2) not spoil: the sequence $\{y_i\}$ must be uniformly distributed

Let $G\colon \mathbb{Z}/2^n \to \mathbb{Z}/2^m$ $(m \le n)$, and let $|G^{-1}(z)|$ does not depend on $z \in \mathbb{Z}/2^m$ (i.e., let $|G^{-1}(z)| = 2^{n-m}$ for each $z \in \mathbb{Z}/2^n$). In cryptography these G are called *balanced*. In case $n - m$ is big, G (seemingly) satisfies both 1) and 2).

Note: In case $n := n \cdot k$ and $m := m \cdot k$ one defines a mapping $G\colon (\mathbb{Z}/2^k)^n \to (\mathbb{Z}/2^k)^m$ to be *balanced modulo* 2^k by analogy.

Theorem 8. *Let $G\colon \mathbb{Z}_p^n \to \mathbb{Z}_p^m$ be a compatible mapping. The function G is balanced modulo p^k for all $k = 1, 2, \ldots$ iff G preserves measure* (i.e., $\mu(G^{-1}(U)) = \mu(U)$ for every measurable $U \subset \mathbb{Z}_p^m$).

Thus, we have to
describe measure-preserving functions among all compatible ones.

Theorem 9. *Let a compatible function $G\colon \mathbb{Z}_p^n \to \mathbb{Z}_p^m$ be uniformly differentiable modulo p on $\mathbb{Z}_p$. Then G preserves measure* $\boxed{\textit{if}}$ *it is balanced modulo p^k for some $k \ge N_1(G)$, and the rank of Jacobi matrix $G_1'(\mathbf{u})$ modulo p is exactly m for all $\mathbf{u} \in (\mathbb{Z}/p^k)^n$.*

Examples. Consider a polynomial $G(x_1, \dots, x_n) \in \mathbb{Z}_p[x_1, \dots, x_n]$

- A polynomial G preserves measure if it is balanced modulo p and all its partial derivatives vanishes simultaneously modulo p at no point of $(\mathbb{Z}/p)^n$.
- In case $p = 2$ one may replace arbitrarily some $+$'s in G with XOR's and/or multiplications by coefficients with AND's; the assertion still remains true.

Note. In case $m = n$ conditions of Theorem 9 are also necessary. For $m = n = 1$ explicit descriptions could be obtained (similar to those of ergodicity). They could be of use also to satisfy conditions 1) and 2) by making the output function G of our PRNG key-dependent (thus making G not known to an adversary).

Theorem 10. (cf. Theorem 4) *For $m = n$ within conditions of Theorem 9 a compatible function G preserves measure* $\boxed{\textit{iff}}$ *it is balanced modulo $p^{N_1(G)+1}$.*

Theorem 11. *For $p = 2$ the function $G\colon \mathbb{Z}_p \to \mathbb{Z}_p$ is compatible and measure-preserving $\Leftrightarrow$*

- (cf. Theorem 2) $G(x) = c_0 + c_1' \cdot x + \sum_{i=1}^{\infty} c_i \cdot p^{\lfloor \log_p i \rfloor + 1} \binom{x}{i}$, *for suitable* $c_i \in \mathbb{Z}_p$, $\|c_1'\|_p = 1$.
- (cf. Theorem 6) $G(x) = c + c' \cdot x + p \cdot U(x)$ *for a compatible function* $U\colon \mathbb{Z}_p \to \mathbb{Z}_p$, $c, c' \in \mathbb{Z}_p$, $\|c'\|_p = 1$.

For $p \neq 2$ only $\Leftarrow$ is true.

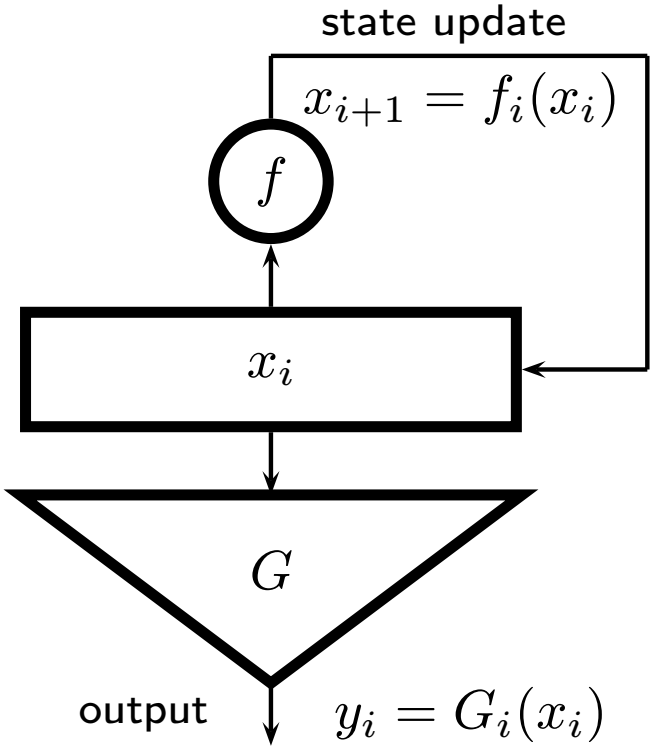
Theorem 12. *Let p be an arbitrary prime.*

- (cf. Theorem 3) *A function $G \in \mathcal{B}$ preserves measure iff G is bijective modulo p^2.*
- (cf. Theorem 5) *A polynomial $G(x) \in \mathbb{Q}_p[x]$ is integer-valued, compatible, and measure-preserving iff the mapping*

$$z \mapsto G(z) \bmod p^{\lfloor \log_p(\deg G) \rfloor + 3}$$

defines a compatible permutation on $\mathbb{Z}/p^{\lfloor \log_p(\deg G) \rfloor + 3}$.

Advantage: This approach leads to flexible PRNGs, where both the state update and output functions are key-dependent. Moreover, we can make both the state update and output functions to be clock-dependent to construct a *counter-dependent automaton.*



A counterpart in dynamics is a *non-autonomous* dynamical system.

To construct these counter-dependent automatons we in fact make use of a *skew shift* $(z, x) \mapsto (U(z), V_z(x))$. To avoid somewhat combersome statements, we restrict ourselves with practical

Examples. Let $c_0, \dots, c_{m-1} \in \mathbb{Z}_2$, $h_0, \dots, h_{m-1}$ be compatible mappings $\mathbb{Z}_2 \to \mathbb{Z}_2$, $0 \leqslant j \leqslant m-1$. The sequence $\{x_i\}$ is periodic modulo 2^k and strictly uniformly distributed modulo 2^k *, and the length of its shortest period is $m \cdot 2^k$, if

1) $m = 2^k$, $\sum_{j=0}^{m-1} c_j \equiv 1 \pmod 2$, and $f_j(x) = c_j + x + 4 \cdot h_j(x)$;

2) $m > 1$ odd, all $h_0, \dots, h_{m-1}$ are ergodic, and
   - $\sum_{j=0}^{m-1} c_j \equiv 0 \pmod 2$,
   - the sequence $\{c_{i \bmod m} \bmod 2 \colon i = 0, 1, 2, \dots\}$ is periodic; m is the length of its shortest period,

   and $f_j(x) = c_j \,\mathsf{XOR}\, h_j(x)$, or $f_j(x) = c_j + h_j(x)$.

*i.e., each $a \in \mathbb{Z}/2^k$ occurs at the period the same number of times

How random is the output? What tests will it provably pass?

Frequency tests are those that consider occurences of (overlapping) ℓ-tuples in a binary output. That is, one represents $x_i \bmod 2^k$ as a k-bit word $\overline{x_i \bmod 2^k}$ (base-2 expansion of $x_i \bmod 2^k$), considers a concatenation

$$\overline{x_i \bmod 2^k}\ \overline{x_{i+1} \bmod 2^k}\ \overline{x_{i+2} \bmod 2^k}\ \ldots$$

and counts occurences of patterns $0, 1, 00, 01, 10, 11, 000, 001, \ldots$. It turnes out that the period of the sequence $\{\overline{x_i \bmod 2^k}\}$ satisfies the following condition

$$\left| \frac{\nu(\beta_0 \ldots \beta_{\ell-1})}{T} - \frac{1}{2^\ell} \right| \le \frac{1}{\sqrt{T}}$$

for all $0 < \ell \le \log_2 T$, where $\nu(\beta_0 \ldots \beta_{\ell-1})$ is the number of occurences of the pattern $\beta_0 \ldots \beta_{\ell-1}$, and T is the length of the period of the sequence $\{\overline{x_i \bmod 2^k}\}$, i.e., $T = mk \cdot 2^k$ in our case.

Linearity tests are those that consider linear dependencies in the output sequence. (To simplfy, return to autonomous systems.)

Definition 3. Let $\mathcal{Z} = \{z_i\}$ be a sequence over a ring R. The *linear complexity* $\lambda_R(\mathcal{Z})$ of $\mathcal{Z}$ over R is the smallest $r \in \mathbb{N}_0$ such that $\exists c, c_0, c_1, \dots, c_{r-1} \in R$ (not all equal to 0) $\forall i = 0, 1, 2, \dots$

$$c + \sum_{j=0}^{r-1} c_j \cdot z_{i+j} = 0. \tag{2}$$

Geometrically: Let $R = \mathbb{Z}/p^k$; all points $(\frac{z_i}{p^k}, \frac{z_{i+1}}{p^k}, \dots, \frac{z_{i+r-1}}{p^k})$, $i = 0, 1, 2, \dots$ fall into parallel hyperplanes.

Proposition 1. *Let* $f(x) \in \mathbb{Q}_p[x]$ (of T. 5) *and let* $\deg f \geq 2$*; then* $\lim_{n\to\infty} \lambda_{\mathbb{Z}/p^n}(\{x_i \bmod p^n\}) = \infty$ (tends not slower than $\log n$)

Note: Let $R = \mathbb{Z}/2$, let $\mathcal{Z}$ be a random sequence of length T. Then the expectation of $\lambda_{\mathbb{Z}/2}(\mathcal{Z})$ is $\frac{T}{2}$.

Proposition 2. $\lambda_{\mathbb{Z}/2}(\{\overline{x_i \bmod 2^n}\}) = 2^{n-1} + 1$.

Polynomial-time tests and 'provable' security. To prove a cipher is secure one makes a 'polynomial-time' reduction to one of plausible (but still unproven) conjectures of 'intractability' of a certain problem, which is 'hard in average'.

Consider a polynomial $\psi(\chi_0, \chi_1, \dots, \chi_{n-1})$ over $\mathbb{Z}/2$ in variables $\chi_0, \chi_1, \dots, \chi_{n-1}$; for $m \in \mathbb{N}$ replace χ_j^m with χ_j. Thus one obtains a *Boolean polynomial*, or an *algebraic normal form* of a Boolean function. To determine whether k Boolean polynomials in n variables have a common zero is an $\mathcal{NP}$-complete problem. We conjecture: For $k \le n$ it is intractable to find a solution of a system of random Boolean equations*. Now we construct a compatible and ergodic function out of given Boolean polynomials ψ_i: For $x \in \mathbb{Z}_2$ denote $\Psi_i(x) = \psi_i(\delta_0(x), \dots, \delta_{n-1}(x)) \in \{0, 1\} \subset \mathbb{Z}_2$; let $\oplus = \mathsf{XOR}$;

$$f(x) = (1 + x) \oplus 2^{n+1} \cdot \Psi_0(x) \oplus 2^{n+2} \cdot \Psi_1(x) \oplus \dots \oplus 2^{n+k} \cdot \Psi_{k-1}(x)$$

*under assumption that the number of monomials in each equation is polynomially restricted

To construct a PRNG we take $f \bmod 2^{n+k+1}$ as a state update function, $G = \lfloor \frac{z}{2^{n+1}} \rfloor \bmod 2^k$ (a truncation of $n+1$ low order bits) as an output function, and $x_0 \in \{0, 1, \ldots, 2^n - 1\}$ as a key. The produced output sequence attains all the above mentioned properties (period of length 2^{n+k+1}, uniform distribution, etc.)

However, it is not difficult to show that to find a state $x = \chi_0 + \chi_1 \cdot 2 + \cdots \chi_{n-1} \cdot 2^{n-1}$ given an output, an adversary (with probability $1 - \frac{1}{2^n}$) has to solve a Boolean system

$$\psi_i(\chi_0, \chi_1, \ldots, \chi_{n-1}) = \varepsilon_i \quad (i = 1, 2, \ldots, k),$$

where $\varepsilon_i \in \{0, 1\}$ are determined by the output.

Moreover, it is possible to construct a counter-dependent automaton (which produces an output sequence that attains all the above mentioned properties) in such a way, that at each new step an adversaray will have to solve a new Boolean system, i.e., the left hand part of a system will change from step to step.

Boolean representation and multivariate mappings. To achieve better performance usually it is better to work with shorter words organized into array of bigger dimension, than with longer words and (say) 1-dimensional array. Thus we need to construct multivariate ergodic (also, measure-preserving) mappings. We construct them *out of univariate ones*. Represent a mapping $F\colon \mathbb{Z}_2 \to \mathbb{Z}_2$ as a correspondence between infinite binary strings

$$x = \chi_0 + \chi_1 \cdot 2 + \chi_2 \cdot 2^2 + \cdots \quad \overset{F}{\mapsto} \psi_0(x) + \psi_1(x) \cdot 2 + \psi_2(x) \cdot 2^2 + \cdots$$

$$(\chi_0, \chi_1, \chi_2, \ldots) \quad \overset{F}{\mapsto} (\psi_0(x); \psi_1(x); \psi_2(x); \ldots)$$

where each $\psi_j(x)$ is a Boolean function of (infinite) number of Boolean variables $\chi_0, \chi_1, \ldots$. It is easy to see that *F is compatible $\Leftrightarrow$ each ψ_j does not depend on $\chi_{j+1}, \chi_{j+2}, \ldots$*. In other words, iff F is a skew shift:

$$(\chi_0, \chi_1, \chi_2, \ldots) \overset{F}{\mapsto} (\psi_0(\chi_0); \psi_1(\chi_0, \chi_1); \psi_2(\chi_0, \chi_1, \chi_2); \ldots).$$

Theorem 13. (A re-statement of a folklore result from the theory of Boolean functions.) *A mapping $F\colon \mathbb{Z}_2 \to \mathbb{Z}_2$ is compatible and measure preserving iff for each $i = 0, 1, \dots$ the Boolean function $\psi_i = \delta_i(F)$ in Boolean variables $\chi_0, \dots, \chi_i$ could be represented as a Boolean polynomial of the form*

$$\psi_i(\chi_0, \dots, \chi_i) = \chi_i + \varphi_i(\chi_0, \dots, \chi_{i-1}),$$

where φ_i is a Boolean polynomial. The mapping F is compatible and ergodic iff, additionaly, the Boolean function φ_i is of odd weight; that is, φ_i takes value 1 exactly at the odd number of points $(\varepsilon_0, \dots, \varepsilon_{i-1})$, where $\varepsilon_j \in \{0, 1\}$ for $j = 0, 1, \dots, i-1$. The latter takes place if and only if $\varphi_0 = 1$, and the degree of the Boolean polynomial φ_i for $i \geq 1$ is exactly i, that is, φ_i contains a monomial $\chi_0 \cdots \chi_{i-1}$.

Now take ergodic

$F\colon (\chi_0, \chi_1, \chi_2, \ldots) \overset{F}{\mapsto} (\psi_0(\chi_0); \psi_1(\chi_0, \chi_1); \psi_2(\chi_0, \chi_1, \chi_2); \ldots)$,

arrange

$$\begin{array}{lllllll}
\chi_0 & \chi_s & \chi_{2s}\ldots & \overset{f_0}{\mapsto} & \psi_0(x) & \psi_s(x) & \psi_{2s}(x)\ldots \\
\chi_1 & \chi_{s+1} & \chi_{2s+1}\ldots & \overset{f_1}{\mapsto} & \psi_1(x) & \psi_{s+1}(x) & \psi_{2s+1}(x)\ldots \\
\ldots & \ldots & \ldots & \ldots & & & \\
\chi_{s-1} & \chi_{2s-1} & \chi_{3s-1}\ldots & \overset{f_{s-1}}{\mapsto} & \psi_{s-1}(x) & \psi_{2s-1}(x) & \psi_{3s-1}(x)\ldots
\end{array}$$

Assuming left strings are variables $x^{(0)}, x^{(1)}, \ldots, x^{(s-1)}$, we conclude that the s-variate mapping

$$\mathbf{F} = (f_0, f_1, \ldots, f_{s-1})\colon \mathbb{Z}_2^s \to \mathbb{Z}_2^s,$$

which is conjugate to a univariate mapping $F\colon \mathbb{Z}_2 \to \mathbb{Z}_2$, is compatible and ergodic.

Example: Let $F(x) = 1 + x$; then $\psi_0 = \chi_0 + 1$, and $\psi_j = \chi_j + \chi_0 \cdots \chi_{j-1}$ for $j \geq 1$.

The corresponding s-variate mapping

$$\mathbf{F}(\mathbf{x}) = (f^{(0)}(\mathbf{x}), \ldots, f^{(s-1)}(\mathbf{x}))$$

is of the following form:

$$f^{(k)}(\mathbf{x}) = x^{(k)} \oplus \left(\left(\bigwedge_{t=0}^{k-1} x^{(t)} \right) \wedge \left(\bigwedge_{r=0}^{s-1} ((x^{(r)} + 1) \oplus x^{(r)}) \right) \right) =$$

$$x^{(k)} \oplus \left(\left(\bigwedge_{t=0}^{k-1} x^{(t)} \right) \wedge \left(\left(\left(\bigwedge_{r=0}^{s-1} x^{(r)} \right) + 1 \right) \oplus \left(\bigwedge_{r=0}^{s-1} x^{(r)} \right) \right) \right),$$

where $\mathbf{x} = (x^{(0)}, \ldots, x^{(s-1)}) \in \mathbb{Z}_2^s$, $k = 0, 1, 2, \ldots, s-1$. (Here and further $\oplus$ stands for XOR.)

Proposition 3. *Let $t, j \in \{0, 1, \ldots, s-1\}$, let all $f_t^{(j)} : \mathbb{Z}_2 \to \mathbb{Z}_2$ be ergodic and let all $g_t^{(j)} : \mathbb{Z}_2 \to \mathbb{Z}_2$ preserve measure. Then the mapping $\mathbf{F}(\mathbf{x}) = (f^{(0)}(\mathbf{x}), \ldots, f^{(s-1)}(\mathbf{x}))$, where*

$$f^{(0)}(\mathbf{x}) = x^{(0)} \boxplus \left(\bigwedge_{r=0}^{s-1} (f_r^{(0)}(x^{(r)}) \oplus x^{(r)}) \right);$$

$$f^{(1)}(\mathbf{x}) = x^{(1)} \boxplus \left(g_0^{(1)}(x^{(0)}) \wedge \left(\bigwedge_{r=0}^{s-1} (f_r^{(1)}(x^{(r)}) \oplus x^{(r)}) \right) \right);$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$f^{(s-1)}(\mathbf{x}) =$$

$$= x^{(s-1)} \boxplus \left(\left(\bigwedge_{t=0}^{s-2} g_t^{(s-1)}(x^{(t)}) \right) \wedge \left(\bigwedge_{r=0}^{s-1} (f_r^{(s-1)}(x^{(r)}) \oplus x^{(r)}) \right) \right),$$

$\mathbf{x} = (x^{(0)}, \ldots, x^{(s-1)}) \in \mathbb{Z}_2^s$, $\boxplus \in \{+, \oplus\}$, is a compatible and ergodic mapping of $\mathbb{Z}_2^s$ onto $\mathbb{Z}_2^s$.