Analysis of Pseudorandom Sequences

Charles Celerier

NCUR 2012

What is a pseudorandom sequence?

What is a pseudorandom sequence?

- uniform distribution

What is a pseudorandom sequence?

- uniform distribution
- low auto-correlation

Stream Ciphers

Stream Ciphers

$$01001110010000110101010101010010 \quad = \quad \text{NCUR}$$

Stream Ciphers

$$
\begin{array}{rlcl}
 & 01001110010000110101010101010010 & = & \text{NCUR} \\
\oplus & 00000101000010010001100100000010 & & \\
\hline
\end{array}
$$

Stream Ciphers

$$
\begin{array}{rlcc}
 & 01001110010000110101010101010010 & = & \mathsf{NCUR} \\
\oplus & 00000101000010010001100100000010 & & \\
\hline
 & 01001011010010100100110001010000 & = & \mathsf{KJLP}
\end{array}
$$

Why use stream ciphers?

- fast
- easy to implement with hardware
- plaintext length is not always known
- near one-time-pad security

My research

- Boolean functions
- 2-adic integers
- pseudorandom sequences
- shift registers

$\mathbb{F}_2$ or "GF two"

XOR	AND
$0 \oplus 0 := 0$	$0 \cdot 0 := 0$
$0 \oplus 1 := 1$	$0 \cdot 1 := 0$
$1 \oplus 0 := 1$	$1 \cdot 0 := 0$
$1 \oplus 1 := 0$	$1 \cdot 1 := 1$

Table: Binary Operations for $\mathbb{F}_2$

$\mathbb{F}_2^n$ or "GF two to the n"

Example

Let $a, b \in \mathbb{F}_2^3$ such that $a = (1, 0, 1)$ and $b = (0, 1, 1)$ then

$$a + b = (1 \oplus 0, 0 \oplus 1, 1 \oplus 1) = (1, 1, 0)$$
$$a \cdot b = 1 \cdot 0 \oplus 0 \cdot 1 \oplus 1 \cdot 1 = 1$$

Fact

$\mathbb{F}_2^n$ *is a vector space.*

Properties of $x \in \mathbb{F}_2^n$

Definition

Let $x, y \in \mathbb{F}_2^n$. Then $wt : \mathbb{F}_2^n \to \mathbb{N} \cup \{0\}$ is defined by

$$wt(x) := \sum_{i=0}^{n-1} x_i$$

and $d : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{N} \cup \{0\}$ is defined by

$$d(x, y) := w(x + y).$$

Then $wt(x)$ is the *Hamming weight* of x and $d(x, y)$ is the *Hamming distance* between x and y.

Some examples

Example

Let $a, b, c \in \mathbb{F}_2^5$ such that

$$a = (0, 1, 1, 0, 1), \; b = (1, 1, 1, 0, 0), \text{ and } c = (0, 0, 1, 1, 0).$$

Then,

$$\begin{array}{ll} wt(a) = 3 & d(a, b) = 2 \\ wt(b) = 3 & d(a, c) = 3 \\ wt(c) = 2 & d(b, c) = 3. \end{array}$$

Boolean functions in $\mathcal{BF}_n$

Definition

Any function f defined such that

$$f : \mathbb{F}_2^n \to \mathbb{F}_2$$

is a *Boolean function*. The set of all Boolean functions on n variables will be denoted by $\mathcal{BF}_n$.

An example

Example

Let $f = x_0 + x_1$.

x_0	x_1	$f(x_0, x_1)$
0	0	0
1	0	1
0	1	1
1	1	0

Table: Truth Table of f

Characters of $\mathbb{F}_2^n$

Definition

A *character* χ of a finite abelian group G is a group homomorphism from G into a multiplicative group of complex numbers.

Fact

$\chi_\lambda(x) := (-1)^{\lambda \cdot x}$ *where* $\lambda, x \in \mathbb{F}_2^n$ *is a* group character *of* $\mathbb{F}_2^n$.

Discrete Fourier Transform

Definition

The *discrete Fourier transform* or DFT of a Boolean function is defined by

$$\mathcal{F}f(\lambda) = \sum_{x \in \mathbb{F}_2^n} f(x)\chi_\lambda(x) \tag{1}$$

Pseudo Boolean Functions

Definition

$\hat{f}(x) := (-1)^{f(x)}$ and $\hat{\mathcal{BF}}_n = \{\hat{f} : f \in \mathcal{BF}_n\}$.

Lemma

The characters of $\mathbb{F}_2^n$ are functions in $\hat{\mathcal{BF}}_n$ and form an orthonormal basis of that set.

Pseudo Boolean Functions

Lemma

$$\hat{f}(x) = \frac{1}{2^{n/2}} \sum_{\lambda \in \mathbb{F}_2^n} c(\lambda)\chi_\lambda(x) \tag{2}$$

where $c(\lambda)$*, the* Fourier coefficients *of* $\hat{f}(x)$ *are given by*

$$c(\lambda) = \frac{1}{2^{n/2}} \mathcal{F}\hat{f}(\lambda). \tag{3}$$

Rothaus' Definition and First Theorem

Definition

If all of the Fourier coefficients of $\hat{f}$ are ± 1 then f is a *bent function*.

Theorem

If f is a bent function on $\mathbb{F}_2^n$, then n is even, $n = 2k$; the degree of f is at most k, except in the case $k = 1$.

Properties of Bent Functions

1. perfectly non-linear
2. $wt(f) = 2^{n-1} \pm 2^{n/2-1}$
3. $\sum_{x \in \mathbb{F}_2^n} f(x) + f(x+a) = 0 \quad \forall a \in \mathbb{F}_2^n$

2-adic integers

What happens when we write positive integers with infinitely many digits?

2-adic integers

What happens when we write positive integers with infinitely many digits?

You get elements of N-adic integer rings!

$\mathbb{Z}_2$

$$
\begin{aligned}
1 &= 1000\cdots \\
2 &= 0100\cdots \\
3 &= 1100\cdots \\
-1 &= 1111\cdots \\
1/3 &= 1101010101\cdots \\
-1/3 &= 1010101010\cdots
\end{aligned}
$$

I am intentionally skipping the details of how to construct rational numbers such as $1/3$. It is only important to know that as long as the denominator is not divisble by 2, it can be done.

Definition

Let $\alpha = (a_n) \in \mathbb{Z}_2 \setminus (0)$. If m is the smallest number in $\mathbb{N} \cup \{0\}$ such that $a_m \not\equiv 0 \pmod{2}^{m+1}$, then the *2-adic valuation* of α is m, or $\log_2(\alpha) = m$. If $\alpha = 0$, then $\log_2(\alpha) = \infty$.

Example

Let $\alpha = 0001011101111 \cdots$. Then $\log_2(\alpha) = 3$.

Boolean Sequences

Definition

Let (a_n) be a sequence. If T is the smallest integer such that $a_i = a_{i+T}$, then the *minimal period* of (a_n) is T.

Definition

Let $f \in \mathcal{BF}_n$ and $v_i \in \mathbb{F}_2^n$ such that $v_i = B^{-1}(i)$ for $0 \le i < 2^n$. Then,

$$seq(f) = (f(v_0), f(v_1), \cdots, f(v_{2^n-1}), f(v_0), \cdots) \tag{4}$$

is a *lexicographical Boolean sequence*.

2-adic Expansion

Definition

Let $f \in \mathcal{BF}_n$ and $v_i \in \mathbb{F}_2^n$ such that $v_i = B^{-1}(i)$ for $0 \leq i < 2^n$. Then,

$$\alpha_f = (f(v_0), f(v_0) + f(v_1) \cdot 2, \cdots, f(v_0) + \cdots + f(v_i) \cdot 2^i, \cdots) \quad (5)$$

where $\alpha_f \in \mathbb{Z}_2$ is called the *2-adic expansion* of f.

Lemma

The digit representation of α_f *is* $seq(f)$.

LFSR

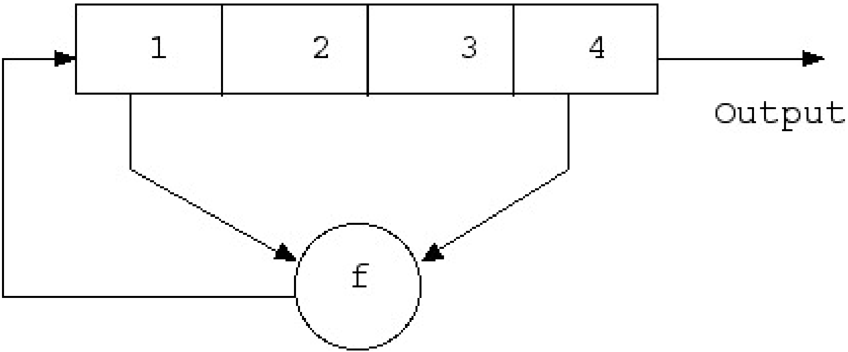

Figure: Linear Feedback Shift Register from Google Images

Eventually periodic shift registers

Solomon W. Golomb wrote the famous book *Shift Register Sequences* in 1967 which contain numerous elementary facts about finite state machines.

Theorem

If the input sequence to a finite state machine is eventually periodic, then the output sequence is eventually periodic.

Breaking a Stream Cipher

Kerckhoffs' principle: "In assessing the security of a cryptosystem, one should always assume the enemy knows the method being used."

Typically, breaking a stream cipher will mean recovering the state of the shift register at a given time.

State of the register

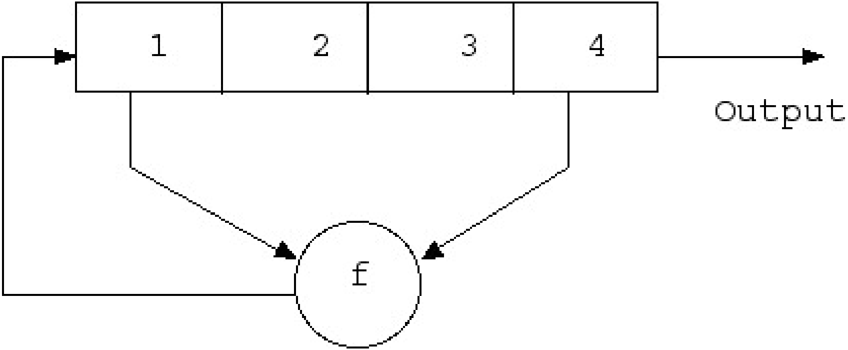

Figure: Linear Feedback Shift Register from Google Images

Two Methods

1. 2-adic integers
2. Boolean sequences

Maiorana-McFarland Class Boolean Functions

A simple bent function construction is accomplished by the Boolean functions in the *Maiorana-McFarland class*. This is the the set $\mathcal{M}$ which contains all Boolean functions on $\mathbb{F}_2^n = \{(x, y) : x, y \in \mathbb{F}_2^{n/2}\}$, of the form:

$$f(x, y) = x \cdot \pi(y) \oplus g(y)$$

where π is any permutation on $\mathbb{F}_2^{n/2}$ and g any Boolean function on $\mathbb{F}_2^{n/2}$. All functions in the Maiorana-McFarland class of Boolean functions are bent.

Using Bent functions for Boolean Sequences

Theorem

The lexicographical Boolean sequence of a Bent function has a period exactly 2^n.

Consider the subset of Maiorana-McFarland class Boolean functions where $g(y) = 0$. $\bar{\pi}$ will be the function which specifies where each index moves to under the permutation π.

Theorem

$\log_2(\alpha_{x \cdot \pi(y)}) = 2^{n/2} + 2^{\bar{\pi}(y_0)}$

The 2-adic valuation of the Boolean sequence of the functions in this subset is entirely dependent on the permutation π.

Conclusion

- Pseudorandom sequences
- Stream Ciphers
- Analysis using Boolean functions and 2-adic integers
- Connections between Bent functions and 2-adic valuation