Composition Analysis of Stream Ciphers Snow 2, ZUC, Grain and Trivium

Guang Gong

Department of Electrical and Computer Engineering
University of Waterloo
CANADA
<http://comsec.uwaterloo.ca/~ggong>

First International Workshop on ZUC Algorithm, Beijing, December 2-3, 2010

Outline

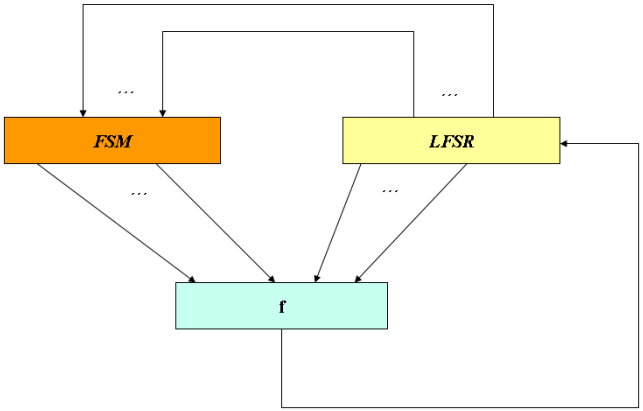
- **General Model** of Pseudorandom Generators Based on FSRs and FSMs
- **Diversity** of Nonlinearity and Differential Distinguisher of the Compositions of Vectorial Boolean Functions
- **Diversity** of Nonlinearity and Differential Distinguisher of ZUC
- **Concluding** Remarks

General Model of Pseudorandom Generators Based on FSRs and FSMs

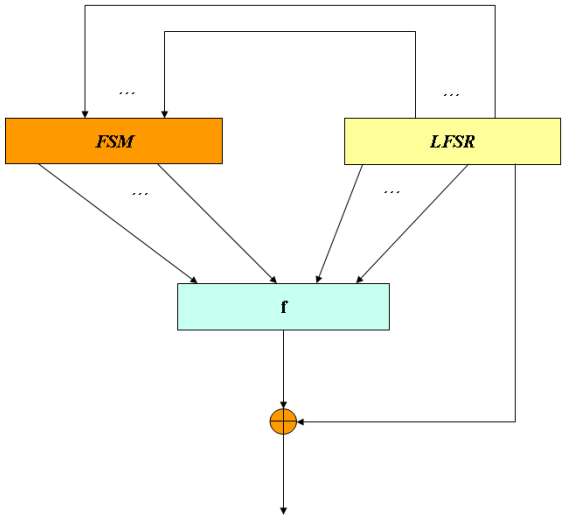
- Due to **algebraic attacks**, it is publicly acknowledged that
  - **filtering** on LFSR (linear feedback shift register) or combinations of LFSRs (e.g., A5), or
  - **simple** combinations of FSM (finite state machine) and LFSR (e.g., E0 in Blue tooth) are not secure

  (Here an LFSR means an *m*-sequence generator, i.e., the feedback polynomial is a primitive polynomial).
- Because of the lack of theory on **nonlinear FSRs**, it is not possible to use nonlinear FSRs alone as pseudorandom generators **(PRGs)**.
- However, **NONE** of the desired randomness properties can be guaranteed without using LFSR in one way or another.
- It turns to design PRGs using an **LFSR** as part of the input of a more sophistic FSM.

General Model

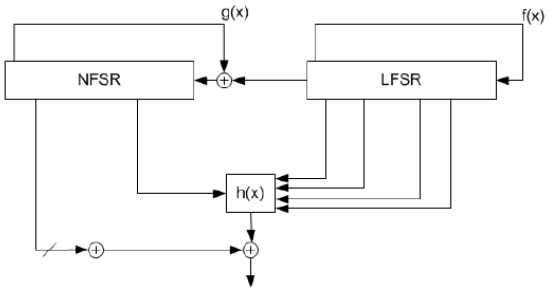


Initialization Phase

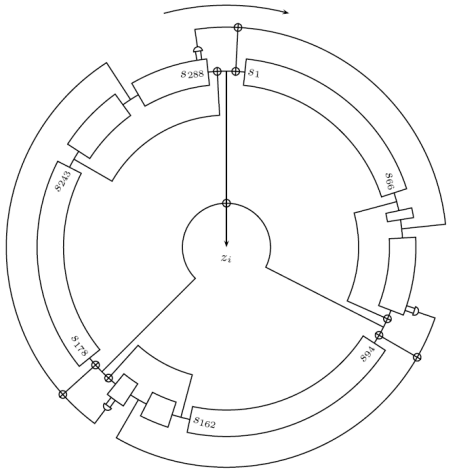


Running Phase

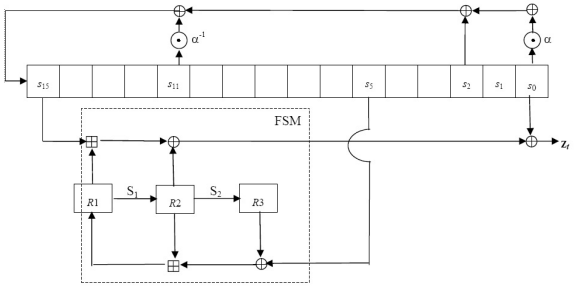
Diagrams of Grain, Trivium, Snow 2 and ZUC from the Literature



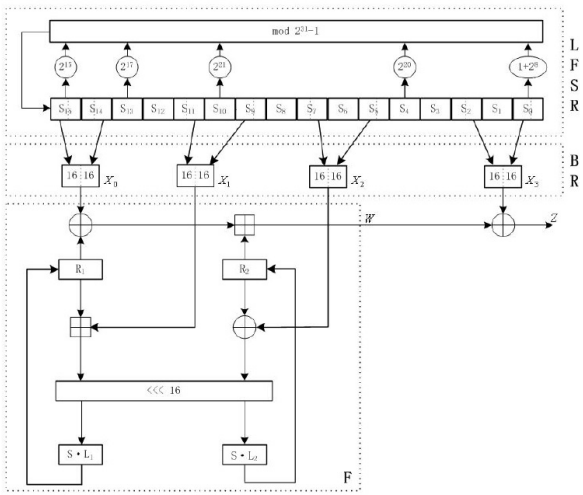
Grain



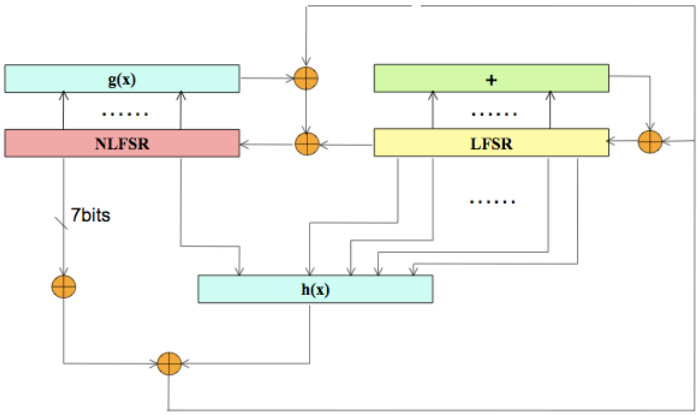
Trivium



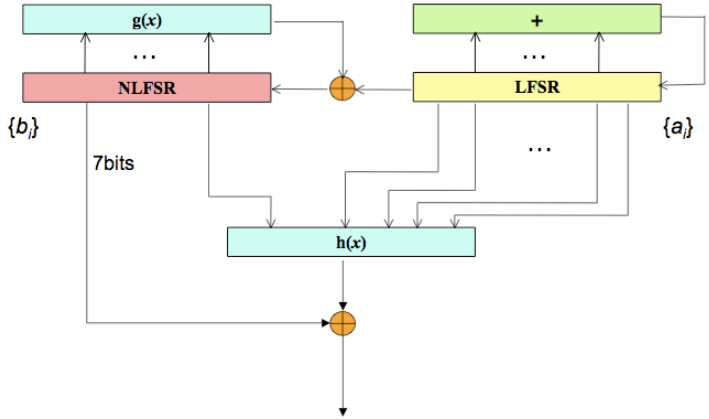
Snow 2



ZUC

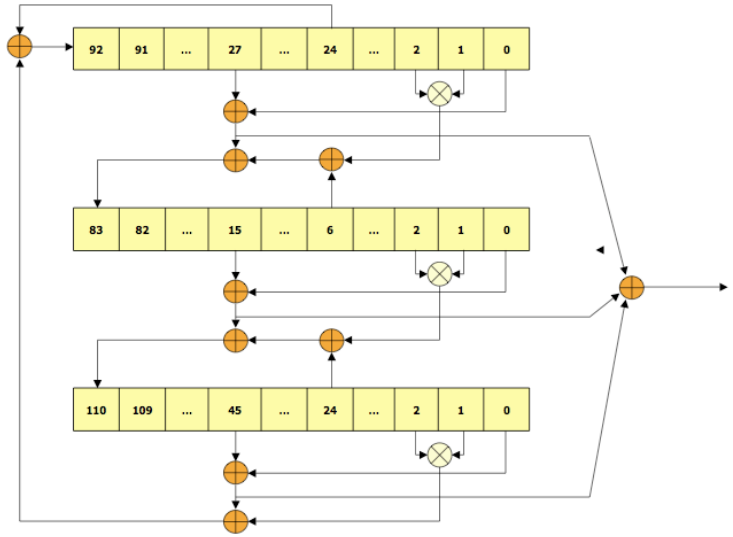


Initialization Phase

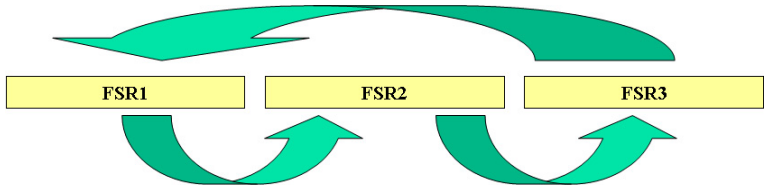


Running Phase

A Diagram of Grain 2 Stream Cipher

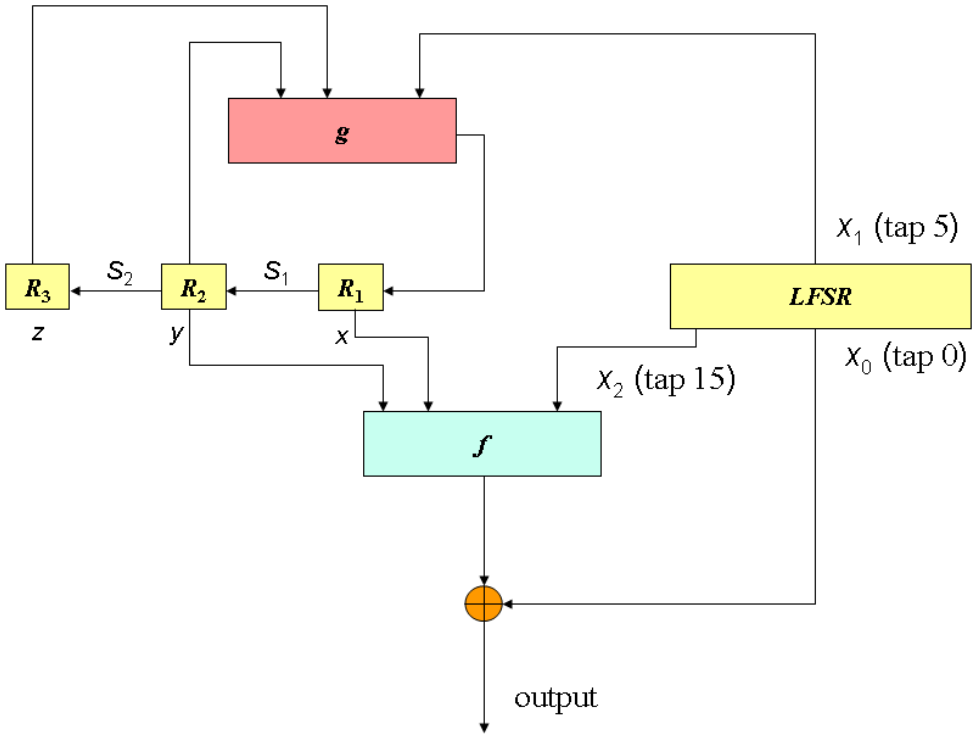


Initialization Phase = Running Phase



A Diagram of Trivium Stream Cipher

Snow 2



Mathematical parameters

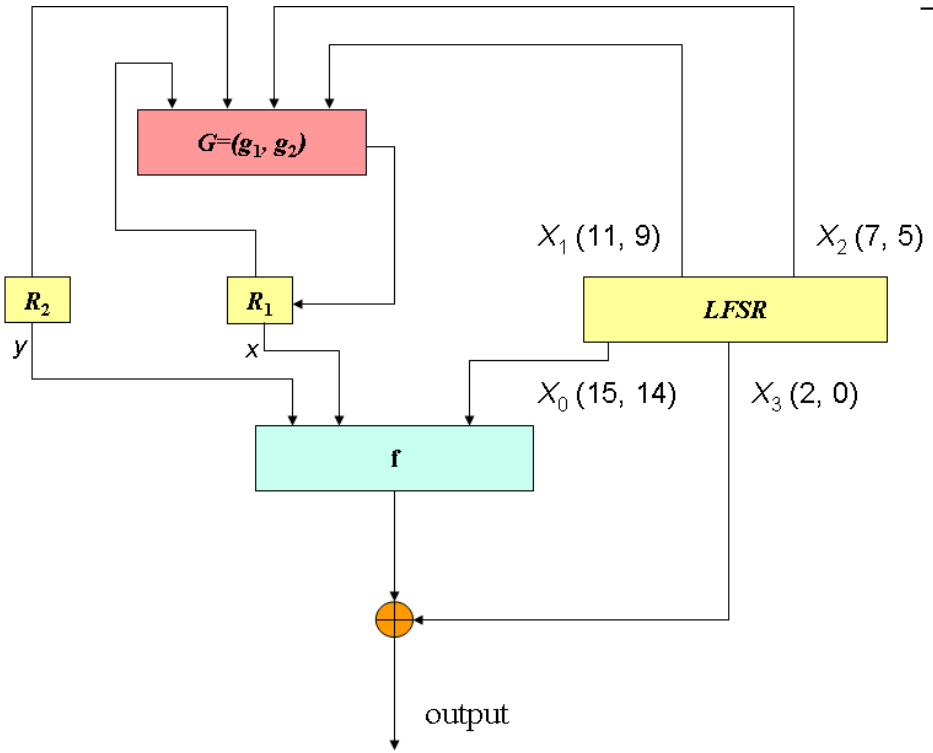
LFSR	degree 16 over $\mathbb{F}_{2^8}$
Output	$f(x, y, x_2) = (x_2 \boxplus x) \oplus y$
FSR	3-stages
$\mathbb{F}_{2^8}$	$= \{\alpha^8+\alpha^6+\alpha^5+\alpha^3+1 = 0\}$
S_Q:	Dickson polynomial, $S_Q(x) = g_{49}(x) + 0x25$, $g_{49}(x) = x \oplus x^9 \oplus x^{13} \oplus x^{15} \oplus x^{33} \oplus x^{41} \oplus x^{45} \oplus x^{47} \oplus x^{49}$
S_R	the Rijndael S-box.

State update function of FSM:

$g(y, z, x_1) = y \boxplus (z \oplus x_1)$ where S_1 and S_2 are computed as follows. Let $\mathbf{w} = (w_0, w_1, w_2, w_3)$, $w_i \in \mathbb{F}_{2^8}$.

$$S_1 = A \cdot \begin{bmatrix} S_Q(w_0) \\ S_Q(w_1) \\ S_Q(w_2) \\ S_Q(w_3) \end{bmatrix}, S_2 = A \cdot \begin{bmatrix} S_R(w_0) \\ S_R(w_1) \\ S_R(w_2) \\ S_R(w_3) \end{bmatrix} \text{ and } A = \begin{pmatrix} \alpha & 1 & 1 & 1+\alpha \\ 1+\alpha & \alpha & 1 & 1 \\ 1 & 1+\alpha & \alpha & 1 \\ 1 & 1 & 1+\alpha & \alpha \end{pmatrix}$$

ZUC



$f(x, y, X_0) = (X_0 \oplus x) \boxplus y$

G is defined as follows.

$u = x \boxplus X_1 = (u_0, u_1, u_2, u_3), u_i \in \mathbb{F}_{2^8}, v = y \oplus X_2 = (v_0, v_1, v_2, v_3), v_i \in \mathbb{F}_{2^8}$.

$L_i : \mathbb{F}_2^{32} \to \mathbb{F}_2^{32}$, linear transforms. Let L be the left cyclic shift operation.

$L_1(x) = x \oplus L^2(x) \oplus L^{10}(x) \oplus L^{18}(x) \oplus L^{24}(x)$

$L_2(x) = x \oplus L^8(x) \oplus L^{14}(x) \oplus L^{22}(x) \oplus L^{30}(x)$

$g_1(x, y, X_1, X_2) = S \circ L_1(u_0, u_1, v_2, v_3)$

$g_2(x, y, X_1, X_2) = S \circ L_2(u_2, u_3, v_0, v_1)$

$S = (S_0, S_1, S_0, S_1)$

S_0 is based on three transformations P_0, P_1, and P_2.

Let $\mathbb{F}_{2^8}$ be defined by $x^8 + x^7 + x^3 + x + 1$. Then $S_1(x) = Mx^{-1} + 0x55$, where M is a binary matrix of order 8×8.

Randomness Properties

- **Trivium**: none, but it provides the proof of a concept that how small of a number of gates in implementation of a cipher could be.
- **Grain**: the period is the multiple of the period of the LFSR (Hu and Gong 09) where the degree of LFSR is the half of the internal state. Balanced.
- **Snow 2 and ZUC**: the period is equal to the period of the LFSR where the degree of LFSR is the number of the internal state. Balanced.

Diversity of Nonlinearity and Differential Distinguisher of Vectorial Boolean Functions

Notation

- $\mathbb{F}_2 = \{0, 1\}$.
- $\mathbb{F}_2^n = \{(x_0, \cdots, x_{n-1}) \mid x_i \in \mathbb{F}_2\}$, a binary vector space of dimension n. We denote $\mathbf{x} = (x_0, \cdots, x_{n-1}), x_i \in \mathbb{F}_2$.
- $\mathbb{F}_{2^n}$, a finite field of 2^n elements.

Vectorial Boolean Functions

- A (n, m) vectorial boolean function F is given by

$$F(x_0, \cdots, x_{n-1}) = (f_0(x_0, \cdots, x_{n-1}), \cdots, f_{m-1}(x_0, \cdots, x_{n-1})) \quad (1)$$

where $f_j(\mathbf{x})$'s are boolean functions of n variables.

Balance Property and Uniform Distribution

- An (n, m) vectorial boolean F is said to be balanced if for any $\mathbf{y} = (y_0, \cdots, y_{m-1}) \in \mathbb{F}_2^m$ the following system of equations

$$\begin{aligned} f_0(\mathbf{x}) &= y_0 \\ f_1(\mathbf{x}) &= y_1 \\ &\vdots \\ f_{m-1}(\mathbf{x}) &= y_{m-1} \end{aligned}$$

  has exactly 2^{n-m} solutions $\mathbf{x}$ in $\mathbb{F}_2^n$.
- It is equivalent to say that the following array is a $(2^n, m)$ orthogonal array:

$$\begin{pmatrix} f_0(\mathbf{x}_0), & \cdots, & f_{m-1}(\mathbf{x}_0)) \\ f_0(\mathbf{x}_1), & \cdots, & f_{m-1}(\mathbf{x}_1)) \\ \vdots & & \\ f_0(\mathbf{x}_{2^n-1}), & \cdots, & f_{m-1}(\mathbf{x}_{2^n-1})) \end{pmatrix}$$

  where $\mathbb{F}_2^n = \{\mathbf{x}_0, \cdots, \mathbf{x}_{2^n-1}\}$.
- In this case, we also say that F satisfies **uniform** distribution.

Nonlinearity

- The Hadamard (or Walsh or Fourier) transform of f:

$$\hat{f}(\mathbf{w}) = \sum_{\mathbf{x} \in \mathbb{F}_2^n} (-1)^{f(\mathbf{x}) + \mathbf{w} \cdot \mathbf{x}}$$

  where $\mathbf{w} = (w_0, \cdots, w_{n-1}) \in \mathbb{F}_2^n$ and $\mathbf{w} \cdot \mathbf{x} = \sum_{i=0}^{n-1} w_i x_i$, the inner product of $\mathbf{w}$ and $\mathbf{x}$.

- The nonlinearity of f, denoted as N_f:

$$N_f = \min_{\mathbf{w} \in \mathbb{F}_2^n, c \in \mathbb{F}_2} d(f, \mathbf{w} \cdot \mathbf{x} + c) \tag{2}$$

  where d is the hamming distance function, or equivalently

$$N_f = 2^{n-1} - \frac{1}{2} \hat{f}_{\max} \tag{3}$$

  where

$$\hat{f}_{\max} = \max_{\mathbf{w} \in \mathbb{F}_2^n} |\widehat{f(\mathbf{w})}|.$$

Nonlinearity of Vectorial Boolean

Nonlinearity of an (n, m) vectorial boolean function $F = (f_0(\mathbf{x}), \cdots, f_{m-1}(\mathbf{x}))$ is defined as

$$N_F = \max_{\mathbf{a} \in \mathbb{F}_{2^m}} N_{\mathbf{a} \cdot F}. \tag{4}$$

Or equivalently,

$$N_F = 2^{n-1} - \frac{1}{2} \hat{F}_{\max} \tag{5}$$

$$\hat{F}_{\max} = \max_{\mathbf{a}, \mathbf{w} \in \mathbb{F}_2^n} |\widehat{\mathbf{a} \cdot F}(\mathbf{w})|$$

which is the maximum correlation between F and affine linear functions.

Diversity of nonlinearity of Vectorial Boolean F

Diversity of nonlinearity of F is defined as the difference between the maximum of the Hadamard spectrum of F and the average value of Hadamard spectra of a random function:

$$Div_F = \left| \frac{\hat{F}_{\max}}{2^n} - \frac{\sqrt{2^n}}{2^n} \right|.$$

We may further write it as

$$Div_F = \left| \frac{\hat{F}_{\max} - \sqrt{2^n}}{2^n} \right| = \left| \frac{c-1}{\sqrt{2^n}} \right|$$

where $\hat{F}_{\max} = \sqrt{2^n}c$ where c is a constant. Note that $1 \leq c < \sqrt{2^n}$ because $\hat{F}_{\max} < 2^n$. Hence, the diversity of nonlinearity of F is to converge to 0 if a vectorial boolean which has a good nonlinearity.

(t, K) Differentials: A New Concept

- A (first order) differential function of boolean function f at $\mathbf{a}$ is defined as
$$f'(\mathbf{a}) = f(\mathbf{x}) + f(\mathbf{x} + \mathbf{a}).$$
- We say that a (n, m) vectorial boolean F has a (t, K)-*differential* where $1 \leq t \leq m$ and $0 \leq k < 2^n$ if for a fixed $\mathbf{a} = (a_0, \cdots, a_{n-1} \in \mathbb{F}_{2^n}$ and for any t-subset: $\{i_1, \cdots, i_t\} \in \mathbb{Z}_m$, the following system of t equations

$$\begin{cases} f'_{i_1}(\mathbf{a}) = f_{i_1}(x_0, \cdots, x_{n-1}) + f_{i_1}(x_0 + a_0, \cdots, x_{n-1} + a_{n-1}) = y_0 \\ \vdots \\ f'_{i_t}(\mathbf{a}) = f_{i_t}(x_0, \cdots, x_{n-1}) f_{i_t}(x_0 + a_0, \cdots, x_{n-1}) + a_{n-1}) = y_{t-1} \end{cases}$$

has at most K solutions $\mathbf{x} \in \mathbb{F}_2^n$ when $\mathbf{y}$ runs through $\mathbb{F}_2^t$.

Differential Distinguishers

- A differential distinguisher of F is defined as the difference between the differential of F and the uniform distributed differential

$$U_{F,t} = \left| \frac{D_{F,t}}{2^n} - \frac{1}{2^t} \right|.$$

  where $2^{n-t}/2^n = 1/2^t$ is the normalized uniformly distributed (t, K) differential of a random function.
- This values is close to zero if a function is indistinguishable from a function with uniform distributed differential.
- Thus, the larger $U_{F,t}$, then easier it is distinguished from the uniform ones.

Cryptographic Properties of the Compositions of Vectorial Boolean Functions

We assume that a function we consider in the paper has zero constant term. Let f and g be two (n, m) vectorial boolean functions and $a(x) = (a_0(x), \cdots, a_{n-1}(x))$ and $b(x) = (b_0(x), \cdots, b_{n-1}(x))$ be two (k, n) vectorial boolean functions. The compositions of f and a and g and b are

$$\begin{aligned} F(x) &= f \circ a(x) = f(a_0(x), \cdots, a_{n-1}(x)) \textbf{ and} \\ G(x) &= g \circ b(x) = g(b_0(x), \cdots, b_{n-1}(x)) \end{aligned} \tag{6}$$

$$\begin{array}{c} \mathbb{F}_2^k \\ \downarrow \quad a(x), b(x) \\ \mathbb{F}_2^n \\ \downarrow \quad f(x), g(x) \\ \mathbb{F}_2^m \end{array} \Longrightarrow \begin{aligned} F &= f \circ a : \\ G &= g \circ b : \end{aligned} \qquad \mathbb{F}_2^k \to \mathbb{F}_2^m$$

Linear Images

For simplicity, we denote $Q = 2^k$, $q = 2^n$, $\mathbb{F}_{2^k} = \{x_0, \cdots, x_{Q-1}\}$ and $\mathbb{F}_{2^n} = \{z_0, \cdots, z_{q-1}\}$. For a given $\lambda \in \mathbb{F}_{2^k}$, we define the following $2^k \times (2n)$ array:

$$M_{a,b}(\lambda) = \begin{pmatrix} a(x_0) & b(\lambda x_0) \\ a(x_1) & b(\lambda x_1) \\ \vdots & \\ a(x_{2^k-1}) & b(\lambda x_{2^k-1}) \end{pmatrix} \tag{7}$$

which is referred to as an image array of a and b with shift λ.

Distribution of Linear Images (Zieler, 1959)

For $k \geq 2n$, if $a(x)$ and $b(x)$ are linear, then they satisfy the following conditions:

1. For $\lambda \notin \mathbb{F}_q$, if $b(\lambda x) \neq \eta b(x)$ for some $\eta \in \mathbb{F}_q$, then the image array of a and b with shift λ is a $(2^k, 2n)$ orthogonal array where **each $2n$-bit vectors occurs 2^{k-2n} times in the array**.
2. For $\lambda \in \mathbb{F}_q$, if there exists some $\eta \in \mathbb{F}_{2^m}$ such that $b(\lambda x) = \eta b(x)$ for all $x \in \mathbb{F}_{2^k}$, then there exists some permutation function $h : \mathbb{F}_q \to \mathbb{F}_q$ such that for any pair $(h(y), \eta y) \in \mathbb{F}_q^2$ occurs 2^{k-n} times in the image array of a and b, which now becomes

$$M_{a,b}(\eta) = \begin{pmatrix} a(x_0) & \eta b(x_0) \\ a(x_1) & \eta b(x_1) \\ \vdots & \\ a(x_{2^k-1}) & \eta b(x_{2^k-1}) \end{pmatrix} \tag{8}$$

Diversity of Nonlinearity of Compositions: Theorem 1

We consider the Hadamard transform of F, i.e., the case G is linear, thus b is linear. If $a(x)$ is linear, then we have the following assertions.

(a) The image array of a and b belongs to the simple image array, and the Hadamard transform of F, is given by

$$\hat{F}(\lambda, \mathbf{u}, \mathbf{v}) = \begin{cases} 0 & \text{if } \lambda \text{ belongs to Case 1} \\ 2^{k-n}\hat{f}(\eta) & \text{if } \lambda \text{ belongs to Case 2} \end{cases}$$

(b) The nonlinearity of F is given by

$$N_F = 2^{k-1} - 2^{k-n-1} N_f \textbf{ or equivalently } \hat{F}_{\max} = 2^{k-n}\hat{f}_{\max}.$$

(c) The diversity of nonlinearity is

$$Div_F = \left| \frac{2^{k/2-n}\hat{f}_{\max} - 1}{\sqrt{2^k}} \right|$$

Differential Distinguishers of Compositions: Theorem 2

If a is linear, then (t, K) differential of F is given by

$$D_{F,t} = 2^{k-n} D_{f,t}$$

and the differential distinguisher of F is given by

$$U_{F,t} = \left| \frac{D_{f,t}}{2^n} - \frac{1}{2^t} \right|, 1 \leq t \leq m.$$

Diversity of Nonlinearity and Differential Distinguisher of S_0 in ZUC

Experimental Results of the S-box S_0

Table: Hadamard Transform: $\widehat{S}_0(\mathbf{u}, \mathbf{v}) = \sum_{x \in \mathbb{F}_{2^8}} (-1)^{\mathbf{u} \cdot S_0(x) + \mathbf{v} \cdot x}$

	Output Components	
	{0, 3, 6}	{ 5, 6, 7}
Maximum Values	64	-64
u	00010010	00000100
v	10000010	00100100
Number of $(\mathbf{u}, \mathbf{v})$	18	28

(t, K)-Differential Properties of S_0

Table: Differentials of S_0 in ZUC

t	$D_{S_0,t}$	$D_{S_1,t}$	Differentials for Uniform Distribution
1	192	144	128
2	160	86	64
3	**128**	**48**	**32**
4	64	28	16
5	40	16	8
6	24	10	4
7	16	6	2
8	8	4	1

From the above table, the worse case is $t = 3$, i.e., $S_0 = (f_0, \cdots, f_7)$ has $(3, 128)$ differentials.

$(3, K)$ Differential of S_0

We investigate this case in details. For $t = 3$, let a 3-subset be $\{i_1, i_2, i_3\}$. Then the following system of equations:

$$\begin{aligned} y_0 &= f_{i_1}(x_0, \cdots, x_7) + f_{i_1}(x_0 + a_0, \cdots, x_7 + a_7) \\ y_2 &= f_{i_2}(x_0, \cdots, x_7) + f_{i_2}(x_0 + a_0, \cdots, x_7 + a_7) \\ y_3 &= f_{i_3}(x_0, \cdots, x_7) + f_{i_3}(x_0 + a_0, \cdots, x_7 + a_7) \end{aligned}$$

have 128 solutions for the following two cases of 3-subsets and the vector $\mathbf{a}$, listed in the following table.

Table: (3, 128) Differentials of S_0

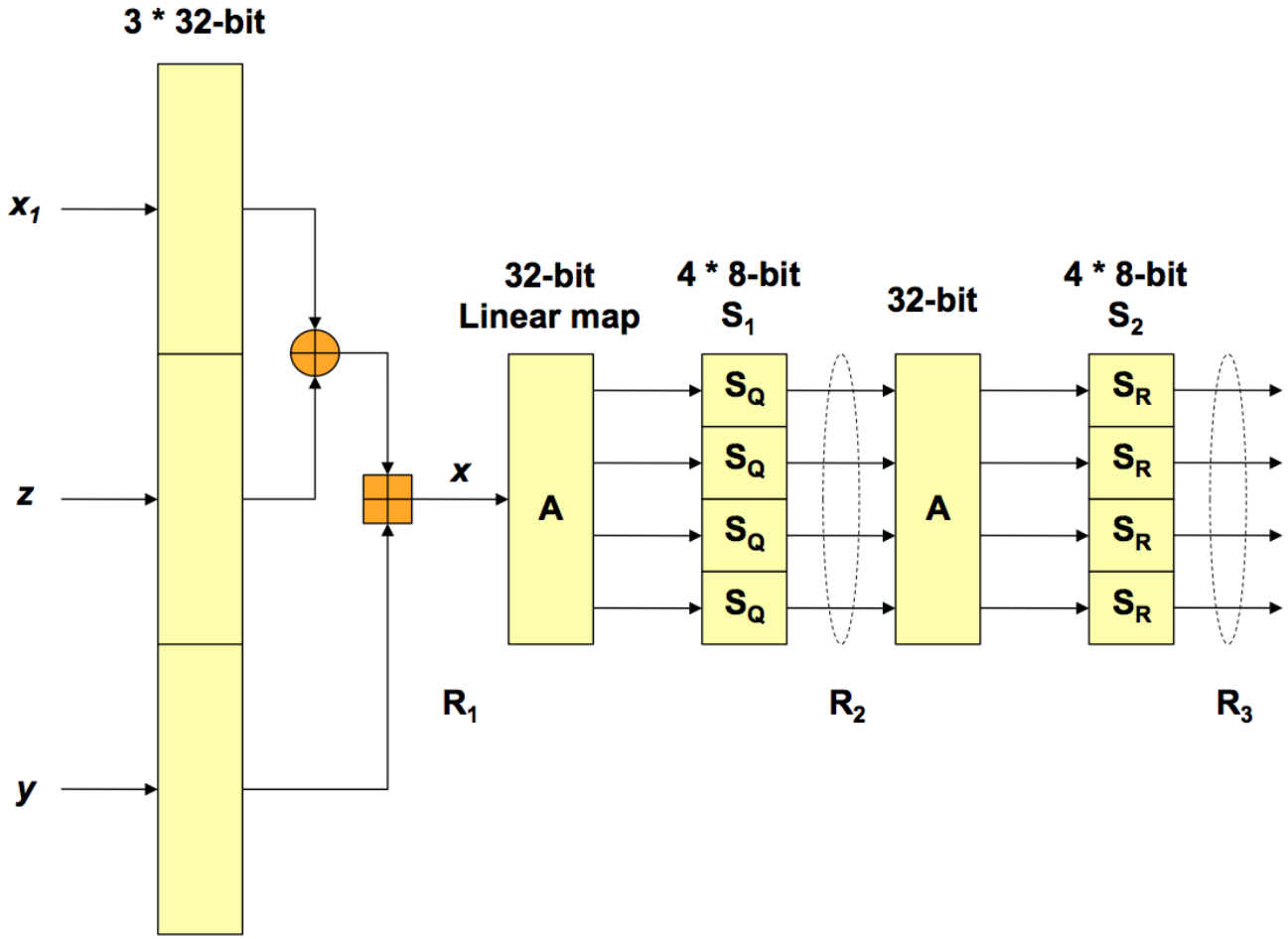
(y_0, y_1, y_2)	(i_1, i_2, i_3)	$\mathbf{a} = (a_0, \cdots, a_7)$
011	(0, 3, 6)	00000101
100	(5, 6, 7)	00000110

Table: Distributions of $(3, 128)$ differentials of S_0

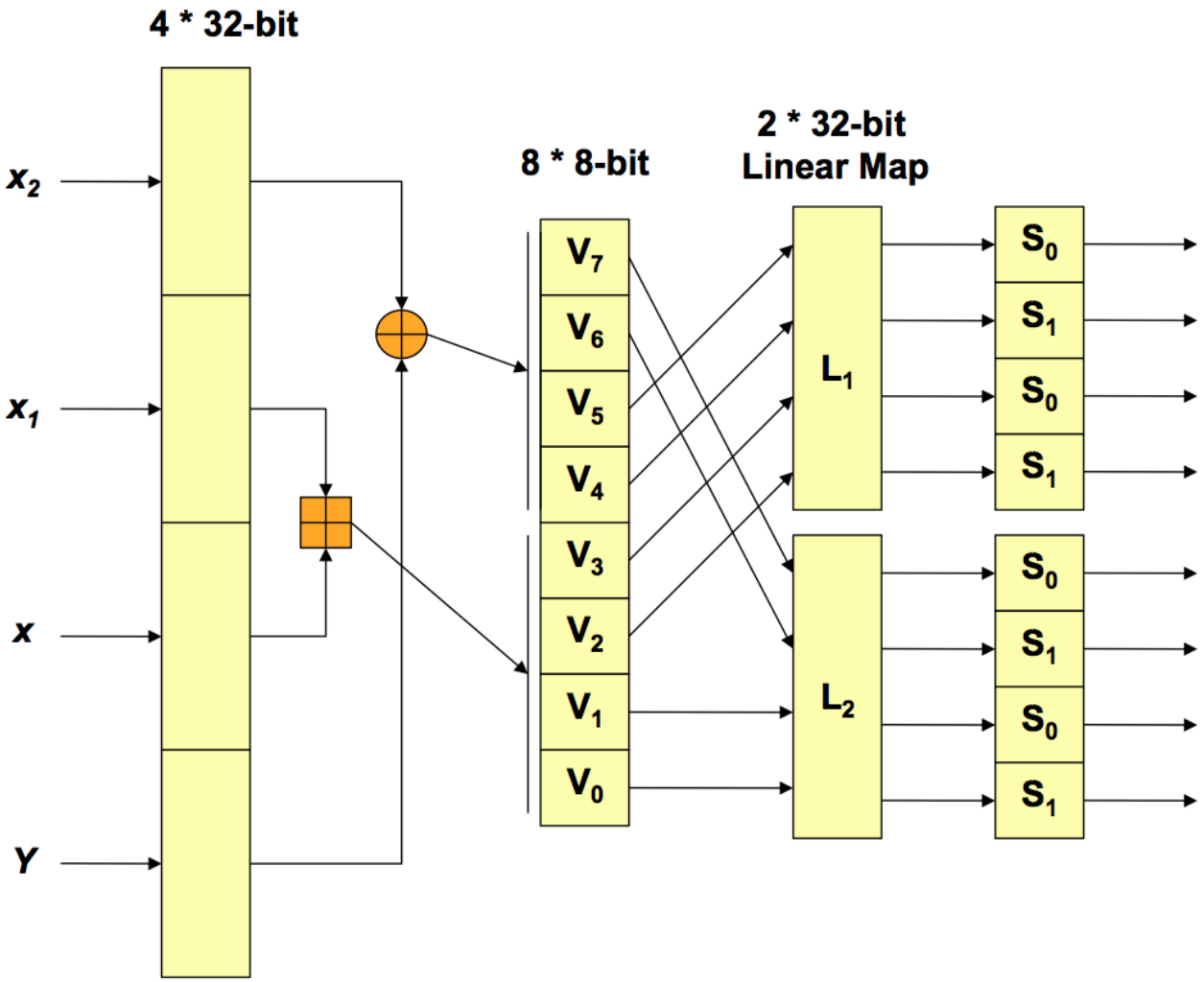
	$D_{S_0,3}$	
(y_0, y_1, y_2)	$\{0, 3, 6\}$	$\{5, 6, 7\}$
000	32	0
100	16	**128**
010	0	32
110	16	32
001	32	32
101	16	32
011	**128**	0
111	16	0

Diversity of Nonlinearity and Differential Distinguisher of ZUC

Compositions of Snow 2



Compositions of ZUC



Assertions

Combining the experiment results on S_0 in ZUC, we have the following results on the diversity of nonlinearity and differential distinguisher of the updating function $S \circ L_i, i = 1, 2$ employed in ZUC, which is the compositions of S_0 and linear functions where $k = 32, n = m = 8$.

Theorem 3

$$Div_{S \circ L_i} = \left| \frac{2^8 \hat{S}_{0max} - 1}{2^{16}} \right| = \left| \frac{2^{14} - 1}{2^{16}} \right| \approx \frac{1}{4} >> 0$$

$$U_{S \circ L_i, 3} = \left| \frac{D_{S_0, 3}}{2^8} - \frac{1}{2^3} \right| = \left| \frac{128}{2^8} - \frac{1}{2^3} \right| = \frac{3}{8} >> 0.$$

Remarks

- Due to the **weakness** of $(3, 128)$ differentials of S_0 in ZUC, i.e., for one 3-bit word, there are 128 solutions to the differential equations, and the rest of seven 3-bit words share the other 128 solutions, the composition of the linear function with S_0 amplifies this weakness largely.
- In other words, if we look at 0th, third and sixth components in $S \circ L_i$, there are a total of $2^{24} \times 128 = 2^{31}$ solutions to the differential equations from those three component functions.
- In this case, the $(3, K)$ differentials of uniform distribution is 2^{29}
- The difference of normalized these two values shows that the differential of $S \circ L_i$ is distinguished from the uniform distributed one, which yields a good differential distinguisher.
- **Counter Measure:** Do not apply any linear transform to input variables of S-boxes, or vectorial booleans.

Conclusions

- **A general structure**: LFSR + FSM where the core part of the state up-dating function of FSM is 8-bit S-boxes (or (8, 8) vectorial boolean functions). Trivium, Grain (80 or 128), Snow 2 and ZUC have a similar key initialization process:

- **Diversity of nonlinearity and differential distinguisher** of the **compositions** of linear functions and nonlinear functions are completely determined by the nonlinear functions.
- **The poor** differential of S_0 in ZUC results in a good differential distinguisher in the state updating function in ZUC, which leads to a distinguish attack.
- This phenomenon shows that the input to an S-box should be not a **linear combination** of multiple inputs.

Acknowledgment

The data on S_0 is computed by Dr. Honggang Hu.