Cryptographically Robust Large Boolean Functions

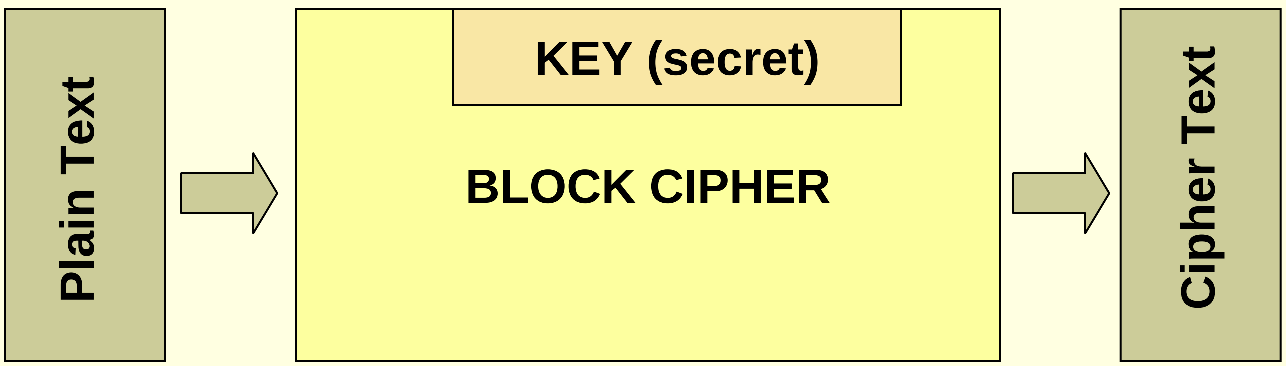
Debdeep Mukhopadhyay
CSE, IIT Kharagpur

Outline of the Talk

- Importance of Boolean functions in Cryptography
- Important Cryptographic properties
- Proposed Construction:
  - Mathematical Formulation
  - Cryptographic Strength
- VLSI Architecture:
  - Scalability
  - Comparisons

Block Ciphers

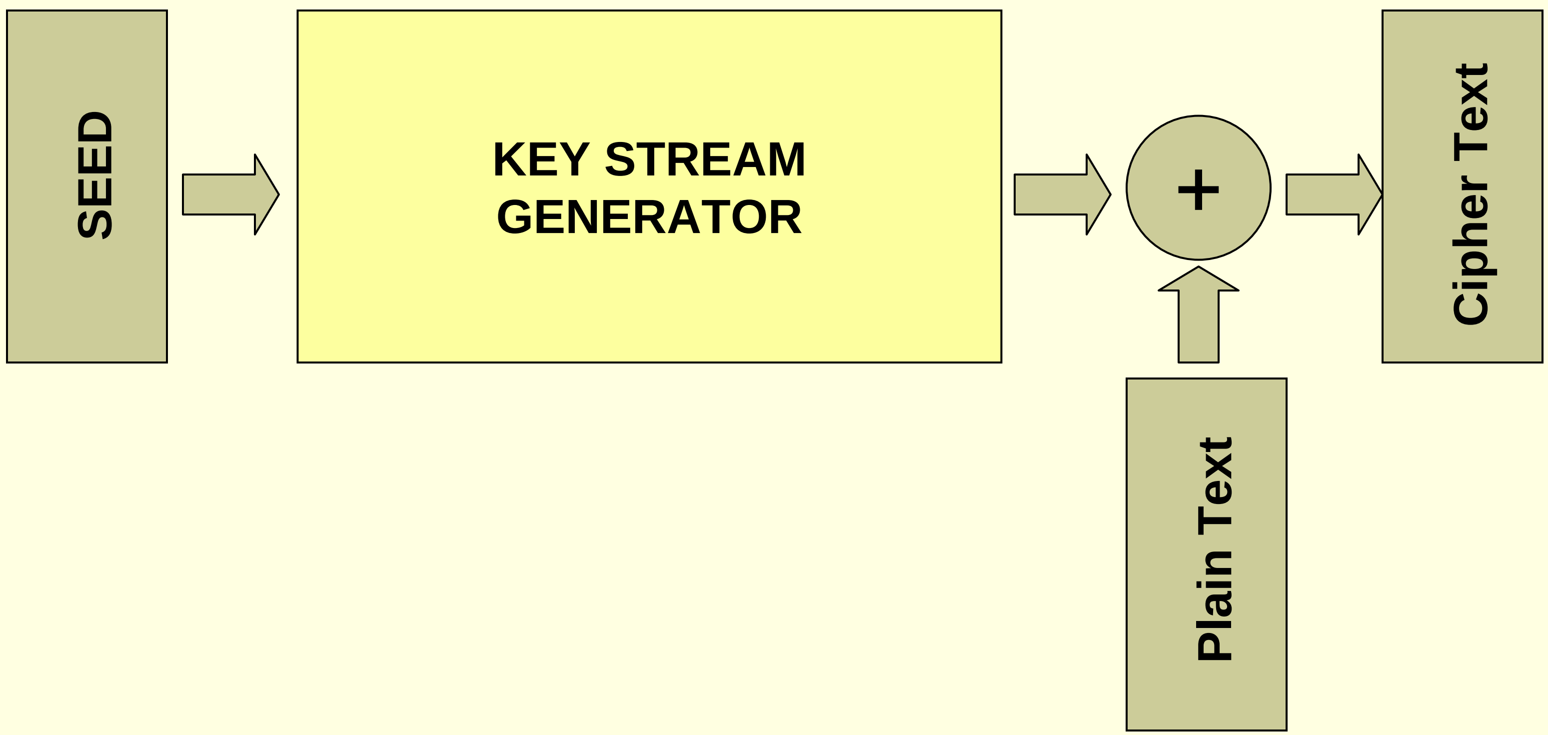
- They are common encryption modules which operate on blocks of data:
  - Ex: AES (Advanced Encryption Standard) operates on 128 bit of data.



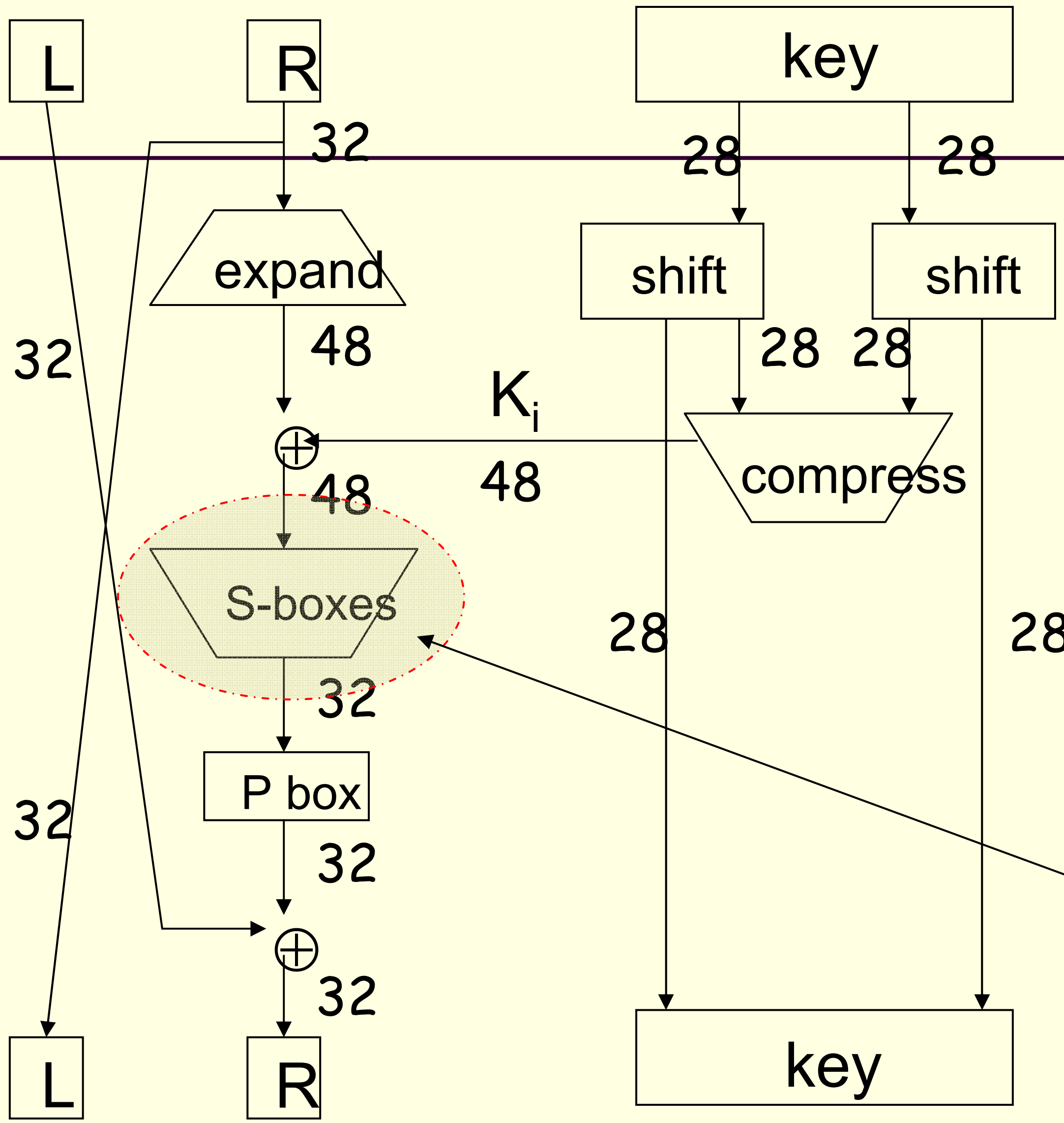
Algorithm is known to the adversary; the key is the only secret

Stream Ciphers

- Typically the block size is 1 or a few bits.



One Round of DES



Q: How to build this?

DES S-box

- 8 “substitution boxes” or S-boxes
- Each S-box maps 6 bits to 4 bits
- S-box number 1

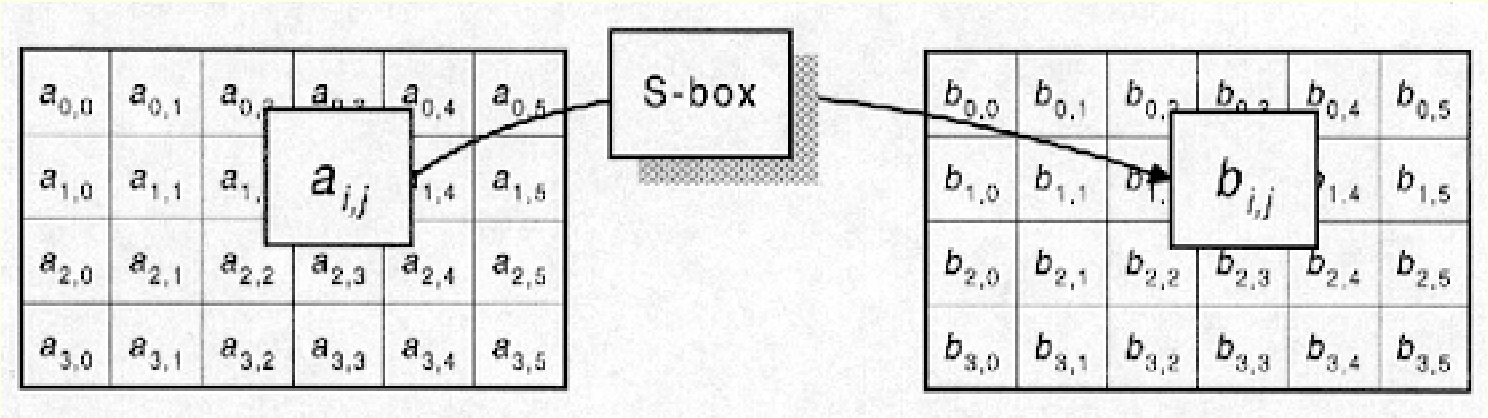
input bits (0,5)
↓
input bits (1,2,3,4)

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
00	1110	0100	1101	0001	0010	1111	1011	1000	0011	1010	0110	1100	0101	1001	0000	0111
01	0000	1111	0111	0100	1110	0010	1101	0001	1010	0110	1100	1011	1001	0101	0011	1000
10	0100	0001	1110	1000	1101	0110	0010	1011	1111	1100	1001	0111	0011	1010	0101	0000
11	1111	1100	1000	0010	0100	1001	0001	0111	0101	1011	0011	1110	1010	0000	0110	1101

What is the design principle?

AES Substitution

- Assume 192 bit block, 4x6 bytes



  - ByteSub is AES's "S-box"
  - Can be viewed as nonlinear (but invertible) composition of some math operations.
  - ***What is the logic behind the construction? What is it based on?***

Design Issues and Modern Challenges

- We require large boolean functions : Typically operating on say 32 bits.
- Area required to implement
  - A Boolean function with n inputs –

<u>Exponential in n</u>

- More complex if we require to generate more than one output simultaneously

Boolean Functions

- Block and stream ciphers can be visualized as Boolean functions.
- A **Boolean function** is a mapping from $\{0,1\}^m \rightarrow \{0,1\}$
- A Boolean function on n-inputs can be represented in minimal sum (XOR +) of products (AND .) form:

$$
\begin{aligned}
f(x_1,\ldots,x_n) = a_0 + a_1.\, x_1 + \ldots + a_n.\, x_n + \\
a_{1,2}.x_1.x_2 + \ldots + a_{n-1,n}.x_{n-1}.x_n + \ldots \\
\ldots + a_{1,2,\ldots,n}\, x_1.x_2 \ldots x_n
\end{aligned}
$$

- This is called the Algebraic Normal Form (ANF)
- If the <u>**and**</u> terms have all zero coefficients we have an affine function
- **If the constant term is further 0, we have a <u>linear</u> function**

Boolean Function

- A Boolean function is a mapping from $\{0,1\}^m \rightarrow \{0,1\}$

$f : \Sigma^n \rightarrow \{0,1\}$ be a Boolean Function.
Binary sequence $(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_{2^n-1}))$
is called the Truth Table of f

- Sequence of a Boolean Function:

$\{(-1)^{f(\alpha_0)}, (-1)^{f(\alpha_1)}, \ldots, (-1)^{f(\alpha_{2^n-1})}\}$ is called sequence of f

Balanced Function

- A Boolean function is said to be balanced if its truth table has equal number of ones and zeros.
- Thus in the sequence of a balanced Boolean function the number of 1s and -1s are the same.

Scalar Product of Sequences

- Consider f and g as two Boolean functions.
- Consider, η be the sequence of f and ε be the sequence of g.
- Define,

$$\boxed{<\eta,\varepsilon>= (\#\text{no of cases when f=g})\text{-}(\#\text{no of cases when f} \neq \text{g})}$$

Non-linearity

- The non-linearity of a Boolean function can be defined as the distance between the function and the set of all affine functions.

$$\therefore N_f = \min_{g \in A_n} d(f, g)$$

where A_n is the set of all affine functions over Σ^n

$$d(f,g) = 2^{n-1} - \frac{1}{2} < \eta, \varepsilon >$$

$$\therefore N_f = 2^{n-1} - \frac{1}{2} \max_{i=0,1,\ldots,2^{n-1}} \{ |\eta, l_i| \},$$

where l_i is the sequence of a linear function in x

A Compact Representation of all the linear functions

- **Hadamard Matrix:** Any rxr matrix with elements in {-1,1} if $HH^T=rI_r$, where I_r is the identity matrix of dimension rxr.
- Walsh Hadamard Matrix:

$$H_0 = 1,\ H_1 = \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix}, n = 1, 2, \ldots$$

- **Each row of H_n is the sequence of a linear function in *x* belonging to $\{0,1\}^n$**
- Each row, l_i is the sequence of the Boolean function,

$g(x) = <\alpha_i, x>, \alpha_i$ is the binary representation of i

Note that α_i and x are not sequences, but they are binary tuples of length n

Balancedness

- The truth-table of the Boolean function has an equal number of 0’s and 1’s.
- XOR is a balanced function.
- AND is an unbalanced function.
- So, we prefer XOR…

Non-linearity

- What is a linear function?
- f is said to be linear wrt + if
  - f(x+y)=f(x)+f(y)

$$x=(x_1,x_2), y=(y_1,y_2), x\oplus y=((x_1\oplus x_2),(y_1\oplus y_2))$$

$$Define,\ f(x)=x_1\oplus x_2.$$

$$\begin{aligned}\therefore f(x\oplus y) &= f(x_1\oplus x_2, y_1\oplus y_2)\\ &= x_1\oplus x_2\oplus y_1\oplus y_2\\ &= f(x)\oplus f(y)\end{aligned}$$

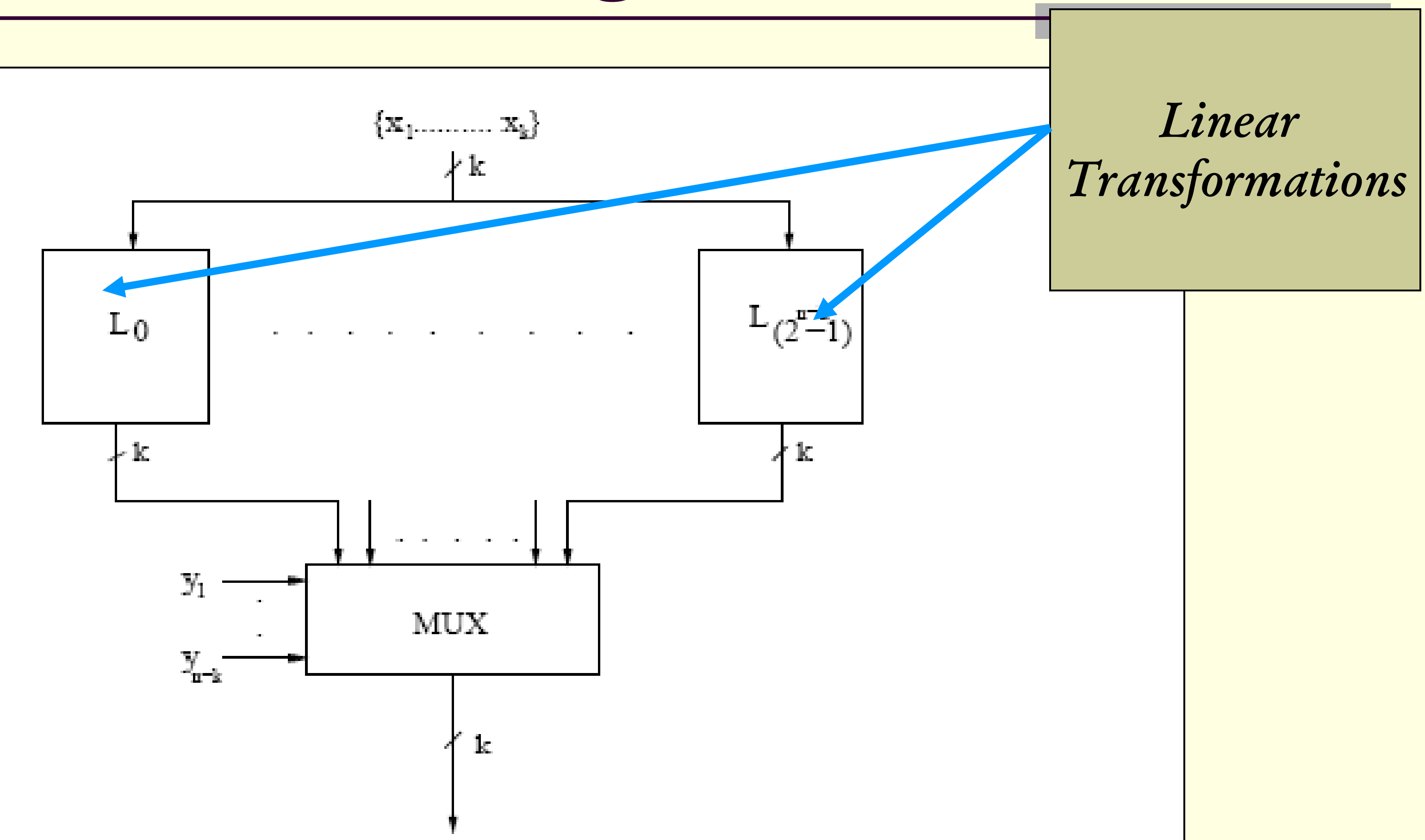
So, XOR is a linear function. But we want non-linear functions.
So, we don't want XOR!

Computing Non-linearity.

x1	x2	x1x2	0	x1	x2	x1^x2
0	0	0	0	0	0	0
0	1	0	0	0	1	1
1	0	0	0	1	0	1
1	1	1	0	1	1	0

Non-linearity is the minimum distance from the truth tables of the linear equations. Here it is 1. So, non-linearity of AND is 1.

A Schematic Diagram



Structure of the Proposed Boolean Function Generator

Construction of $n \times k$ Mapping

- The n-bit input is split into two portions:
  - x of size k bits
  - y of size n-k bits

Input, z = (y, x)

- $2^{(n-k)}$ kxk Linear Transformations
  - Each transformation operates on **x**
  - Converts the k-bit input to a k-bit output
- The multiplexer chooses one of the k bits depending on y

Output, $Q(z) = \{q_1(z), \ldots\ldots, q_k(z)\}$

Properties of the set S

- Formed of linear transformation of order k and elements in GF(2) i,e {0, 1}
- The transformations represented in the form of matrices, T_k have maximal period:

$\forall v \in GF(2^k) \setminus \{0\}$ and $1 \leq i < j \leq 2^k - 1$,

$T_k^i(v) \neq T_k^j(v) \Rightarrow T_k^{2^k-1} = I,$ where I is a $k \times k$

identity matrix.

Properties of the set S

$S = \{I, T_k, \ldots, T_k^{2^k - 2}\}$ contains a set of $2^k - 1$ invertible matrices of dimension $k \times k$

From this set we choose 2^{n-k} linear transformations for the linear array of transformations.

$2^k - 1 > 2^{(n-k)}$

$\leftrightarrow 2^k > 2^{(n-k)}$

$\leftrightarrow k > n-k$

$\leftrightarrow k > n/2$

Properties of the set S

Lemma 1: The transformation $T_k^{i-1} \in S$ is invertible $(1 \leq i < 2^k)$.

Lemma 2: Set S is closed under addition modulo 2.

Lemma 3: If $T_k^{i-1}, T_k^{j-1} \in S$, rows of the matrices T_k^{i-1} and T_k^{j-1} are pairwise distinct when $i \neq j$.

Mathematical Formulation

- Linear transformations can be represented as k x k matrices:

$$Li = \begin{pmatrix} li1 \\ \dots \\ lik \end{pmatrix}, 0 \le i \le 2^{n-k} - 1$$

- Mathematically, the output k-bit vector Q(z) is

$$Q(z) = \oplus_{\sigma=0}^{2^{n-k}-1} D_\sigma(y) L_\sigma(x)$$

$$D_\sigma(y) = (\overline{i_1} \oplus y_1)(\overline{i_2} \oplus y_2)\dots(\overline{i_{n-k}} \oplus y_{n-k}),$$

$$\sigma = (i_1 i_2 \dots i_{n-k}), y = (y_1 y_2 \dots y_{n-k})$$

Cryptographic Properties

Theorem 4: : The non-linearity of each component function $q_i(z)$ $(1 \leq i \leq k)$ is at least $2^{n-1} - 2^{k-1}$, where $k > n/2$.

Theorem 5: : Any component function $q_i(z)$ $(1 \leq i \leq k)$ is balanced.

Theorem 6: : The non-linearity of any non-zero linear combination of the component functions $q_i(z)$ $(1 \leq i \leq k)$ is at least $2^{n-1} - 2^{k-1}$ $(k > n/2)$. The resulting functions are always balanced.

Resiliency

Definition 6: Resiliency: The Boolean function $f(x_1, \ldots, x_n)$ of an n-variable is called correlation-immune of order m $(1 \leq m \leq n-1)$ iff, for any $1 \leq i_1 \leq \ldots \leq i_m \leq n$ and $a_1, \ldots, a_m$, $P(f(X_1, \ldots, X_n) = 1 | X_{i_1} = a_1, \ldots, X_{i_m} = a_m)) = P(f(X_1, ..., X_n) = 1)$ where the X_is are independent and uniformly distributed binary random variables and $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$, $a_i = 0$ or 1; $P(.)$ and $P(.|.)$ mean the probability and the conditional probability respectively [32]. Balanced m^{th} order correlation immune functions are called m-resilient functions [31]

Cryptographic Properties

Theorem 8: : The mapping $Q(z) = \{q_1(z), \ldots, q_k(z)\}$ is a regular mapping from V_n to V_k.

Theorem 9: : The algebraic degree of each component functions of the $n \times k$ mapping $(k > n/2)$ and their non-zero linear combinations is $(n - k + 1)$.

Theorem 10: : The maximum resiliency of the component functions of the $n \times k$ mapping $(k > n/2)$ and their non-zero linear combinations is $k - 2$.

Cryptographic Properties

- For each component function $q_i(z)$
  - Non – linearity is at least $2^{n-1} - 2^{k-1}$, $k>n/2$
  - It is balanced
    - ***Same is true for any non-zero linear combinations***
  - Algebraic degree is at least (n-k+1)
  - Mapping $Q(z) = \{ q_1(z), \ldots\ldots , q_k(z) \}$ is regular from V_n to V_k
- Number of mappings generated is $\boxed{P_{2^{n-k}}^{2^k-1}}$

Strict Avalanche Criterion

- Boolean function f on Vn satisfies SAC iff

  $f(x) \oplus f(x \oplus \alpha)$ is balanced for all $\alpha \in Vn$
- Original construction Q(z) does not satisfy SAC
- For z' = Wz,
  - Q(Wz) satisfies SAC
  - W is a non-degenerate n x n matrix with entries from GF(2)

$$W = \begin{pmatrix} I_{n-k} & 0 \\ D_{kXn-k} & I_k \end{pmatrix}; D = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Example : 8x5 mapping

- **n=8, k>4=5**

$$T = \begin{matrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{matrix}$$

Compute Q(156), assume key=0

$$W=\begin{pmatrix}1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\1&0&0&1&0&0&0&0\\1&0&0&0&1&0&0&0\\1&0&0&0&0&1&0&0\\1&0&0&0&0&0&1&0\\1&0&0&0&0&0&0&1\end{pmatrix}, Z=\begin{pmatrix}1\\0\\0\\1\\1\\1\\0\\0\end{pmatrix}$$

$$Z'=WZ=\begin{pmatrix}1\\0\\0\\0\\0\\0\\1\\1\end{pmatrix}$$

y=4

x=3

z=156

$$Q(156)=T^4(3)=((T^2)^2)(3)$$

$$=\begin{pmatrix}1&1&0&0&1\\1&1&1&1&0\\0&1&1&0&0\\0&1&0&0&0\\1&0&0&0&0\end{pmatrix}\begin{pmatrix}0\\0\\0\\1\\1\end{pmatrix}$$

$$=\begin{pmatrix}1\\1\\0\\0\\0\end{pmatrix}$$

Q(156)=192

Cryptographic Properties

- Non-linearity is 112 which is very high (maximum for 8 variables 120)
- Degree of each function is 4
- All non-zero combinations are balanced and have non-linearity of 112.
- Robustness against Differential Cryptanalysis is 0.848, bias in the Linear Approximation Table is 16.
- Each boolean function satisfies SAC

VLSI Design of the Architecture

- Input y denotes the CA to be selected
  - NB: All the CA are the same machine in different states of evolution (the clock cycles are different)
  - y determines the number of cycles, s, the CA is to be applied
  - A mapping, *g*, from y to s is required=> $Q(z)=T^{g(y)}(x)$
    - (Alternate expression of the construction)
- Domain of g is V_{n-k}, while range is V_k
- One to many mapping (as, k>n/2)
  - *No deterministic hardware possible*

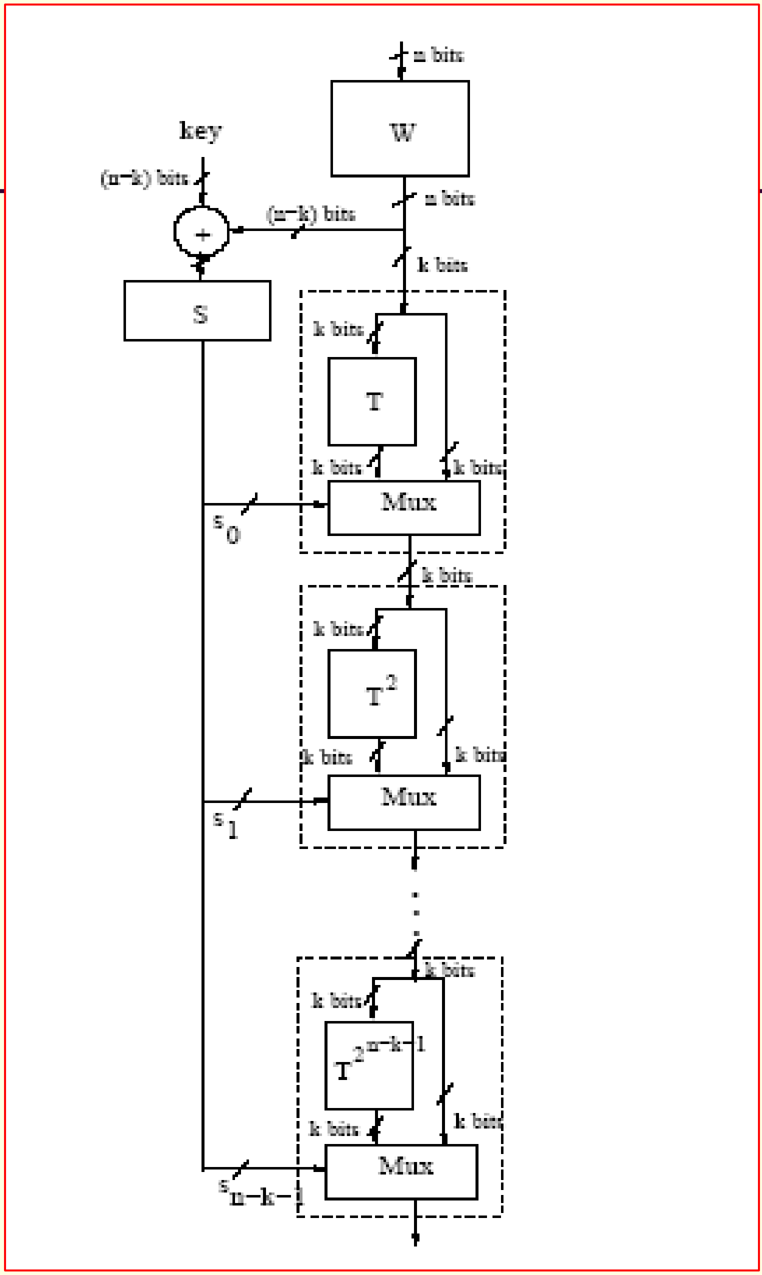
Restricted Design Architecture

- Restrict the clock cycles to $2^{(n-k)}$
- Mapping becomes (n-k) to (n-k)
- Permutation is done by using XORing with a secret k, s
- Value of s for a given y, will depend on the secret key, *key* of n-k bits
- Number of possible permutations 2^{n-k}
- Cryptographic properties remain the same, as this is an equivalent representation.

Restricted Design Architecture

- Each CA is to be cycled s times i.e. T needs to be multiplied s times
- Square and multiply algorithm is used for better performance
- Output is obtained in O(n-k) time

Block Diagram



Experimental Results

Dimension	XOR	MUX	Flip-Flop	Time (clk cycles)
8 x 5	26	15	3	3
10 x 6	54	24	4	4
16 x 9	208	63	7	7
24 x 13	691	141	11	11

Observation: Growth of the resources is polynomial with dimension

Scalability

$$\mu_1 = Hardware/bits$$
$$\mu_2 = (Hardware \times delay)/bits$$

The growth of parameters μ_1 and μ_2 with n and k

n	k	μ_1	μ_2
8	5	11.2	33.6
10	6	25.5	68
16	9	37	260
20	11	54	488
22	12	54	542
26	14	86	1030
30	16	111.5	1561
38	20	175.5	3159
50	26	304	7294
56	29	378	10206

n: No of Input bits
k: No of Output bits
μ_1: $Hardware/bits$
μ_2: $(Hardware \times delay)/bits$

Comparisons

Comparison of a 24 variable Boolean function with that of [3]

Parameter	Method in [3] (Pipelined)	Proposed Method (Pipelined)
Number of Boolean Functions	1	13
Non-linearity	2^{23}	2^{23}
Degree	5	12
Resiliency	18	11
Flip Flops	2196	143
Gates	129	830
Delay (clock cycles)	14	11
μ_1	2325	75
μ_2	32550	823

References

1. D. Mukhopadhyay and D. Roy Chowdhury, “A Parallel Efficient Architecture for Large Cryptographically Robust n x k (k>n/2) Mappings, To Appear in IEEE Transactions on Computers.

2. D. Mukhopadhyay, “Design and Analysis of Cellular Automata based Cryptographic Algorithms”, PhD Thesis, IIT Kharagpur, 2007.

3. Palash Sarkar and Subhamoy Mitra, “Efficient Implementation of Cryptographically Useful "Large" Boolean Functions", IEEE Transactions on Computers, vol. 52, no. 4, pp. 410--416, 2003.

4. K. C. Gupta and P. Sarkar, ``Improved construction of non-linear resilient s-boxes'', IEEE Transactions on Information Theory, vol. 51, no. 1, pp. 339--347, January 2005.

References

5. Jennifer Seberry, Xian-Mo Zhang and Yuliang Zheng, “Systematic Generation of Cryptographically Robust S-boxes,", 1st Conference Computer and Communication Security, VA, USA, 1993, pp. 171--182.

6. K. C. Gupta and P. Sarkar, “Improved Construction of Non-linear Resilent S-Boxes", In Advances in Cryptology-Asiacrypt. 2002, pp. 466--483, Springer Verlag.

7. K. C. Gupta and P. Sarkar, “Efficient Representation and Software Implementation of Resilient Maiorana-McFarland and S-Boxes", In WISA 2004. 2004, pp. 317--331, LNCS 3325, Springer Verlag.

8. K.Nyberg, ``Perfect non-linear S-boxes,'' Advances in Cryptology-Eurocrypt, 1991, pp. 378--386.

References

9. J. Seberry, X. M. Zhang and Y. Zheng, “On Construction and Nonlinearity of Correlation Immune Boolean Functions'', Proceedings of Eurocrypt. 1993, pp. 181--199, Springer Verlag.

10. X. .M. Zhang and Y.Zheng, ``On cryptographically resilient functions'', IEEE Transactions on Information Theory, vol. 43, no. 5, pp. 1740--1747, 1997.

11. E Pasalic and S..Maitra, ``Linear codes in generalised construction of resilient functions with very higly nonlinearity,'' IEEE Transactions on Information Theory, vol. 48, no. 8, pp. 2182--2191, 2002.

12. J. Detombe and S. Tavares, “Constructing Large Cryptographically Strong S-Boxes''. Advances in Cryptology, Crypto, 1992, pp. 165--181.

Small and compact designs survive…

Thank You