KNOWLEDGE-BASED SCHEMA FOR S-BOX DESIGN

Gabriela Moise

Petroleum-Gas University of Ploiesti, no. 39 Blvd. Bucuresti, Ploiesti, Romania

ABSTRACT

In cryptography, a Substitution box (S-box) is one of the basic components of a symmetric key cryptography. Generally, it transforms a number of *m* input bits into *n* output bits. In this paper, there are presented the cryptographic properties of S-boxes and it is introduced a knowledge-based schema for building S-boxes.

Keywords: *Cryptography, S-box design, Strict-strong avalanche criteria, Avalanche effect*

1. INTRODUCTION

Shannon establishes the main important criteria, which define a good cryptosystem, in the paper entitled *Communication Theory of Secrecy Systems* [1]. These criteria are the following: amount of secrecy, size of keys, complexity of enciphering and deciphering operations, propagation of errors, and expansion of message. The perfect secrecy "is defined by requiring of a system that after a cryptogram is intercepted by the enemy the a posteriori probabilities of this cryptogram representing various messages be identically the same as the a priori probabilities of the same messages before the interception." [1]

The substitution tables (S-boxes) play a fundamental role in the block encryption algorithms in order to meet the definition of a perfect secrecy. The generation of these S-boxes has its roots in mathematical functions (Boolean functions). The design of S-boxes considers the concept of perfect secrecy, the simplicity of using the algorithm, and the implementation cost.

Briefly, S-boxes may be defined as nonlinear transformations which map a string of *m* bits to a string of *n* bits.

In order to resist to all possible cryptographic attacks, S-Boxes have to satisfy a set of criteria. The types of possible cryptographic attacks are the following: known plaintext (linear, correlation, algebraic), chosen plaintext (differential), adaptive chosen plaintext, ciphertext-based (known ciphertext, chosen ciphertext, adaptive chosen chiphertext), and encryption key-based attacks.

Researchers have proposed different methodologies to build S-boxes based on the above-mentioned criteria. Some of the methods are presented below:

- Using random S-box-based algorithms, and then testing the obtained S-boxes against the cryptographic criteria;
- The recursive method proposed in [5] enables users to build good *(n+1)*x*(n+1)* S-boxes starting from proper *n*x*n* S-boxes;
- Method presented in [6] uses a bent functions construction and the concept of dynamic distance;
- Algorithms based on heuristic techniques [7];
- Methods based on chaotic maps [8].

In this paper the author proposes a knowledge based schema to build S-boxes.

Starting from the properties of the Boolean functions and from the related theorems, a set of rules which enables the creation of S-Boxes starting from the knowledge is built.

The facts are good S-boxes with small dimension. Once a new proper S-box is discovered, it is added to the knowledge base. More complex and proper S-boxes can be generated by combining proper S-boxes using the generation rules. The rules that compose the rule base are derived from the theorems presented by Lloyd in [9], Kim *et al*. in [5], and methods proposed by Bardis *et al*. in [12]. S-boxes are in order to check if they satisfy the cryptographic properties using techniques presented in [10]. In order to classify S-boxes, there is defined a measure of the "power" of S-boxes depending on their properties.

2. CRYPTOGRAPHIC PROPERTIES OF S-BOXES

In order to work with S-boxes, it is considered the following definition:

A S-box is a Boolean function from Z_2^n to Z_2^m, with $n > m$

$$S : Z_2^n \to Z_2^m \quad (1)$$

Any Boolean function from Z_2^n to Z_2^m, with $n > m$ can be expressed as m Boolean function: $(S_m, S_{m-1}, S_{m-2}, \ldots, S_1)$, where $S_i : Z_2^n \to Z_2$, $Z_2 = \{0,1\}$ and Z_2^n is the array with n binary elements. In the rest the of paper, it is referred as the Boolean function $S : Z_2^n \to Z_2$.

The set of criteria that are used to build proper S-boxes is as follows:

- **A criterion of non-linearity** is needed for the cryptographic algorithm to resist to linear attacks. S-boxes should have high nonlinearity.
- **Strict Avalanche Criteria (SAC)** is necessary to resist to deferential cryptanalysis. The concept was introduced by Webster and Tavares in [3] and this means if complementing a single input bit results in changing the output bit with one half probability.
- **High Order Strict Avalanche Criteria (HOSAC)** is an extension of SAC realized by Forré [4] considering the subfunctions obtained by keeping one or more input bits constant.
- **Bit Independence Criterion (BIC)** states that inverting one single input bit *i*, the output bits *j* and *k* should change independently, for all *i, j, k* [3].

3. PRELIMINARIES

The Walsh transformation of $S(x)$ is an integer, defined as $W_S(\omega) = \sum_{x \in \{0,1\}^n} (-1)^{S(x) + x \bullet \omega}$, where

$$x \bullet \omega = x_1\omega_1 \oplus \ldots \oplus x_n\omega_n. \quad (2)$$

The nonlinearity of S is given by $nl(S) = \frac{1}{2}\left(2^n - \max_{\omega \in \{0,1\}^n} |W_S(\omega)|\right)$. The nonlinearity is maximized when

$$\max_{\omega \in \{0,1\}^n} |W_S(\omega)| \text{ is minim.} \quad (3)$$

A Boolean function is balanced, if $|x \in Z_2^n, S(x) = 0| = |x \in Z_2^n, S(x) = 1|$. (4)

In a different way, a Boolean function is balanced if and only if $W_S(0) = 0$. (5)

A boolean function $S : Z_2^n \to Z_2^m$ is complete if

$$\sum_{x \in Z_2^n} S(x) \oplus S(x \oplus c_i) > (0,0,\ldots,0), \forall i = \overline{1,n}\ ,\ c_i = (0,0,\ldots,1,\ldots 0), \text{with 1 on position } i. \quad (6)$$

The Hamming weight of a boolean function is noted with w and represents the number of 1 in its true table.

In [9], there are presented the following definitions, which are used in the knowledge-based schema for building S-boxes.

$S : Z_2^n \to Z_2^m$ respects the *avalanche effect if* and only if

$$\sum_{x \in Z_2^n} w(S(x) \oplus S(x \oplus c_i)) = m * 2^{n-1}, \forall i = \overline{1,n}\ ,\ c_i = (0,0,\ldots,1,\ldots 0), 1 \text{ on position } i. \quad (7)$$

If $S : Z_2^n \to Z_2^m$, then S satisfies the *strict avalanche criteria* if

$$\sum_{x \in Z_2^n} S(x) \oplus S(x \oplus c_i) = (2^{n-1}, \ldots, 2^{n-1}), \forall i = \overline{1,n}. \quad (8)$$

A Boolean function with n variables satisfies the *SAC of order* k, where $0 \le k \le n-2$, if fixing k inputs bits, the resulting functions satisfy SAC properties and the Boolean fuction satisfies the SAC of order $k-1$.

An order relation can be defined on the set of S-boxes $S : Z_2^n \to Z_2^m$. The default properties for any S-boxes have to be:

- Nonlinear
- Avalanche effect
- Completeness

$S_1 \lhd S_2$, and we call that S_2 is higher than S_1 if S_2 satisfies the SAC of order l and S_1 satisfies the SAC of order p and $l > p$.

Thus, the following relations can be established:

S, with SAC of order 1 $\lhd S$, with SAC of order 2 $\lhd \ldots \lhd S$, with SAC of order n-2. (9)

Observation: In the case of $m=1$: if S satisfies the avalanche effect then S is complete and the SAC criteria is exactly the avalanche effect criteria.

We say that a function satisfies the *strong avalanche effect* if and only if a bit input changing implies changing of average p output bits, where $\frac{m}{2} < p \le m$.

If $S: Z_2^n \to Z_2^m$, then S satisfies the *strict-strong avalanche criteria* (*SOAC*), if

$$\sum_{x \in Z_2^n} S(x) \oplus S(x \oplus c_i) > \left(2^{n-1}, \ldots, 2^{n-1}\right), \forall i = \overline{1,n}. \quad (10)$$

Analoguely, the following order relation can be defined:

$S_1 \prec S_2$. We consider that S_2 is strong-higher than S_1 if S_2 satisfies the SOAC of order l and S_1 satisfies the SOAC of order p and $l > p$.

S, with SOAC of order 1 $\prec S$, with SOAC of order 2 $\prec \ldots \prec S$, with SOAC of order n-2. (11)

4. KNOWLEDGE-BASED SCHEMA FOR S-BOX DESIGN

The schema for S-box design is presented in figure no. 1.

The user states a query which is processed by the inference engine. The inference engine refers knowledge from two fact bases: one is *S-boxes base* and the *other S-boxes Generating Rules base*.

Both bases contain classified knowledge. S-boxes are classified according to the order relation defined in the previous chapter. The rules are classified according to the complexity of the algorithms used in their codification and the types of S-boxes generated.

The procedures of S-boxes classification use the analysis of the hamming weight according to the bit position [10] in order to verify the SAC properties and the hamming weight according to the number of changed bits to verify the SOAC properties.

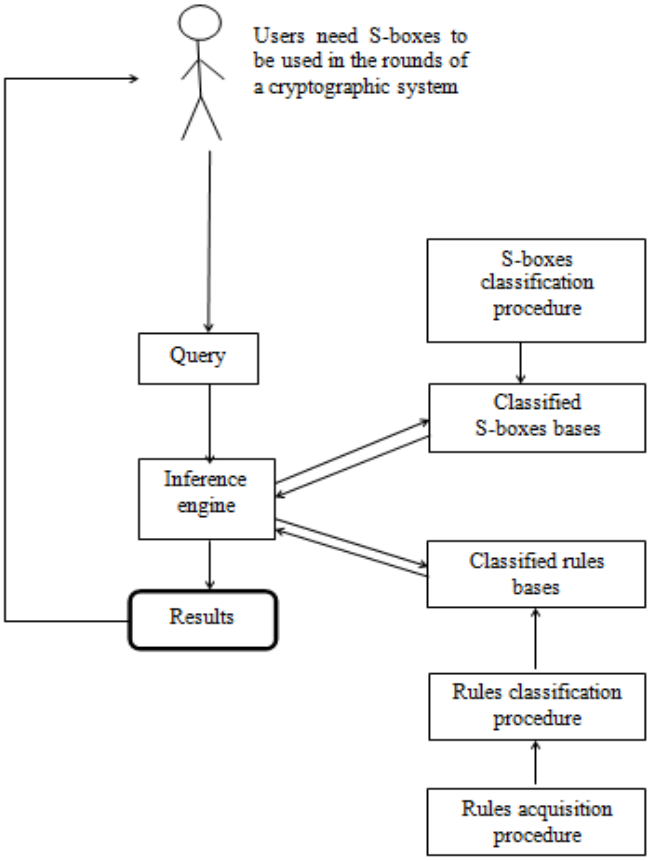

Figure 1 Knowledge-based Schema for S-box design.

In order to define rules that generate proper S-boxes, which will be included in the rules bases, it is used the following theorem demonstrated in [5]:

"If both a bijection $f: Z_2^n \to Z_2^n$ and a function $g: Z_2^n \to Z_2$ satisfy SAC, then for any integer $k \in \{1, 2, \ldots, n\}$, the function $E^k[g, f]: Z_2^{n+1} \to Z_2^{n+1}$ is a bijection satisfying the SAC."

where $E^k[g, f]: Z_2^{n+1} \to Z_2^{n+1}$ is defined as $E^k[g, f](y) = \left(D_1^k[g](y), D_0^k[f_n](y), D_0^k[f_{n-1}](y), \ldots, D_0^k[f_1](y)\right)$, $\forall\, y \in Z_2^{n+1}$ and

$$D_j^k[f](0, x) = f(x),\ D_j^k[f](1, x) = f(x \oplus c_i) \oplus j\ . \quad (12)$$

Starting from this theorem, there can be established n rules such as:

R: If (f is bijection having SAC property and g has the SAC property,) then ($E^k[g, f]$ has the SAC property).

Other rules can be established starting from the methods presented in [12]:

R: If ($f(x_1, \ldots, x_n)$ is balanced SAC function), then ($f(x_1, \ldots, x_j \oplus x_{n+1}, \ldots, x_n), j = \overline{1, n}$ is balanced SAC functions).

Two original methods are presented in [12]: one based on the orthogonal transformations and the other based on combinatorial approach. These techniques are more practical and can be easily implemented in software applications.
So, other rules can be derived from [12]:

R: rule based on orthogonal transformations
R: rule based on combinatorial approach

In [9], Lloyd states that the number of function $f: Z_2^n \to Z_2^n, n \geq 2$ with SAC property of order $n-2$ is 2^{n+1} and provides a method to generate these types of functions. By using them, there are obtained 16 functions satisfying SAC of order 1, for n=3.
In [11], the authors determine formulas to calculate the bounds of balanced SAC functions: for n=2 the number of SAC functions is 8, for n=3 the number is 64, for n=4 the number is 4128, for n=5 the number is 27522560.
One may notice that the number of SAC functions increase quickly. Starting from the above-mentioned statements, there can be derived several rules which can be included in the S-boxes generating rules bases.

5. CONCLUSIONS

In this paper, the author introduced a new property for the Boolean function with cryptographic properties, namely the strong avalanche effect if and only if a bit input changing implies changing of average p output bits, where $\frac{m}{2} < p \leq m$.
The ideal S-box is the function with the property: a bit input changing implies changing of all output bits.
Also, it was delivered a knowledge-based schema to build S-boxes necessary to increase the security level of the cryptographic systems.
The advantages of using such a knowledge based schema are as follows: 1) obtaining large S-boxes; and 2) the possibility to extract S-boxes and to change the cryptographic system whenever necessary.
The future research directions will be focused on the SOAC property and the implementing a software programme in order to build S-boxes.

6. REFERENCES

[1]. C. E. Shannon, Communication Theory of Secrecy Systems, *Bell Systems Technical Journal*, Vol. **28**, pp. 656-715, (1948).

[2]. O. Cangea, Moise, G., A New Approach of the Cryptographic Attacks, *DICTAP2011 Proceedings*, to appear, (2011).

[3]. A.F. Webster and S.E. Tavares. On the design of S-boxes. *Advances in Cryptology : Proc. of CRYPTO'85*, Springer-Verlag, 523–534, (1986).

[4]. R. Forré, The strict avalanche criterion: spectral properties of boolean functions and an extended definition, *Proceeding of CRYPTO '88*, Springer-Verlag, 450-468, (1988).

[5]. K. Kim, T. Matsumoto, H. A. Imai, A Recursive Construction Method of S-boxes Satisfying Strict Avalanche Criterion, *Proceeding of CRYPTO '90*, 564 - 574, (1990).

[6]. Adams, Carlisle M. (Ottawa, CA), Serge, Mister. J. M. (Amherstview, CA) Practical S box design United States Entrust Technologies, Ltd. (Ottawa, CA) 6031911 http://www.freepatentsonline.com/6031911.html, (2000).

[7]. A. Lineham, Heuristic S-box Design, *Contemporary Engineering Sciences*, Vol. **1**, no. 4, 147 – 168, (2008).

[8]. G. Tanga, Xiaofeng Liao, and Yong Chen, A novel method for designing S-boxes based on chaotic maps, *Chaos, Solitons & Fractals* Volume **23**, Issue **2**, 413-419, (2005).

[9]. S. Lloyd, Counting functions satisfying a high order strict avalanche criterion, *Advances in Cryptology — EUROCRYPT '89 Lecture Notes in Computer Science*, **434**, (1990).

[10]. Phyu Phyu Mar, Khin Maung Latt, New Analysis Methods on Strict Avalanche Criterion of SBoxes, *World Academy of Science, Engineering and Technology* **48**, (2008).

[11]. A. M. Youssef, T. W. Cusick, P. Stanica, S.E Tavares, New Bounds on the Number of Functions Satisfying the Strict Avalanche Criterion, *In Third Annual Workshop on Selected Areas in Cryptography*, (1996).

[12]. N. G. Bardis, A. P. Markovskyy, M. Mitrouli, A. Polymenopoulos, Methods for Design of Balanced Boolean Functions Satisfying Strict Avalanche Criterion (SAC), www.wseas.us/e-library/conferences/athens2004/papers/487-825.pdf, accessed at the 30th may (2011).