"1st International Symposium on Computing in Informatics and Mathematics (ISCIM 2011)"
in Collabaration between EPOKA University and "Aleksandër Moisiu" University of Durrës
on June 2-4 2011, Tirana-Durres, ALBANIA

Cryptography, SPN algorithm and a simulation on S-Boxes

Blerina Çeliku, Msc.Dhori Beta, Erma Pema

Department of Natyral Sciences, Fan.S.Noli University

Email: blerina.celiku@yahoo.com
Email: dhoribeta@yahoo.com
Email: ermikaela@yahoo.com

Abstract

It is impossible to imagine ourselves not involved in this tremendously increasing technology nowadays. In the field of networking, the area of network security consists of the provisions and policies adopted by the network administrator to prevent and monitor unauthorized access, misuse, modification, or denial of the computer network and resources that are network-accessible. Network security is a complicated subject, historically only discussed by well-trained and experienced experts. However, as more and more we become ``wired", all of us need to understand the basics of security in a networked world. In this research paper we treat shortly the network concepts such as authentication and anti-virus softs and management of security which is different for all kinds of networks. So a small home or a workplace would only require basic security while large businesses will require high maintenance and advanced software and hardware to prevent malicious attacks from hacking and spamming. The term network security and information security are often used interchangeably. Information security uses cryptography to transform usable information into a form that renders it unusable by anyone other than an authorized user; this process is called encryption. Some history of cryptography is included, as well as an introduction that begins with Caesar Cipher and other used simple cryptosystems such as Shift, Substitution, Affine, Vigenere, Hill, Permutation, and Autokey Ciphers and the classification of modern cryptographic algorithms. In next session will be discussed S-Boxes as the most important part of SPN (Substitution-Permutation Network) that is a mixed transformation of bits in a number of rounds. What is more important is the non-linearity of the s-box within a cryptosystem. That means that the function between the input bits and output bits must be non-linear. This gives the strengthness of an algorithm against attacks. The question is; how can this be achieved? There are several criterias that an s-box must satisfy to be non-linear. These will be described in an example then we go on with a software that examines the properties of the s-box. As a result will be defined if the s-box is appropriate to be used or not. What us really matters is that what must be secret has to be a secret!

Keywords: *Information security, Network, Cryptography, Cipher, S-Box.*

1. Introduction

Network security starts from authenticating the user, commonly with a username and a password. Since this requires just one thing besides the user name, i.e. the password which is something you 'know', this is sometimes termed one factor authentication. With two factor authentication something you 'have' is also used (e.g. a security token or 'dongle', an ATM card, or your mobile phone), or with three factor authentication something you 'are' is also used (e.g. a fingerprint or retinal scan). Communication between two hosts using a network could be encrypted to maintain privacy.
Since the early days of writing, heads of state and military commanders understood that it was necessary to provide some mechanism to protect the confidentiality of written correspondence and to have some means of detecting tampering. Julius Caesar is credited with the invention of the Caesar cipher ca. 50 B.C., which was created in order to prevent his secret messages from being read should a message fall into the wrong hands. World War II brought about many advancements in information security and marked the beginning of the professional field of information security. Also there are several ciphers used in coding theory such as Shift, Substitution, Affine, Vigenere, Hill, Permutation, and Autokey Ciphers. The end of the 20th century and early years of the 21st century saw rapid advancements in telecommunications, computing hardware and software, and data encryption.

2. Cryptography

Cryptography is the science of writing in secret code and is an ancient art; the first documented use of cryptography in writing dates back to circa 1900 B.C. when an Egyptian scribe used non-standard hieroglyphs in an inscription. Cryptography is used in information security to protect information from unauthorized disclosure while the information is in transit and while information is in storage. Cryptography provides information security with other useful applications as well including improved authentication methods, message digests, digital signatures, non-repudiation, and encrypted network communications. Wireless communications can be encrypted using protocols such as WPA/WPA2. Wired communications are secured using AES for encryption and X.1035 for authentication and key exchange. Software applications such as GnuPG or PGP can be used to encrypt data files and Email.
Cryptography can introduce security problems when it is not implemented correctly. The length and strength of the encryption key is also an important consideration. A key that is weak or too short will produce weak encryption. They must be protected from unauthorized disclosure and destruction and they must be available when needed. PKI solutions address many of the problems that surround key management.

The most important issue is to learn the concepts behind basic cryptographic methods, and then to compare the myriad cryptographic schemes that are in use today.
There are several ways of classifying cryptographic algorithms. The cryptographic algorithms based on the number of keys that are employed for encryption and decryption are categorized as:

- Secret Key Cryptography (SKC): Uses a single key for both encryption and decryption as DES, AES, CAST-128/256, IDEA, Rivest Ciphers, Blowfish, Twofish algorithms etc**.**
- Public Key Cryptography (PKC): Uses one key for encryption and another for decryption as: RSA, Diffie-Hellman, DSA, ElGamal algorithms etc.
- Hash Functions: Uses a mathematical transformation to irreversibly "encrypt" information

3. The S-Box Properties

3.1 In 1948 and 1949, the two principles known as "confusion" and "diffusion" have been introduced by Claude Shanon. The first one, describes the undefinite relation between plaintext and ciphertext. The second one, spreads the redundancy of plaintext over the ciphertext. The simplest way of having diffusion is permutation or a "mixed transformation". This transformation is achieved using the SPN (Substitution-Permutation Network). An SPN with a key *k* seems like a map, where bits change their places $f_k: \{0,1\}^N \rightarrow \{0,1\}^N$, where N is the number of plaintexts and ciphertexts. These are vectors of N bits:

$$P = [p_1 p_2 ... p_N], p_i \in \{0,1\} \text{ (plaintext bits) and } C = [c_1 c_2 ... c_N], c_i \in \{0,1\}$$

(ciphertext bits).

In SPN we have R rounds, each one of N bits. In every round we have a substitution and a linear transformation. The S-box nxm is like a map where n and m are integers: $S : \{0,1\}^n \rightarrow \{0,1\}^m$. The number of input bits (n) is equal to number of output bits (m). The number of s-boxes used in every round is equal to M in equation (1),

$$M=N/n \qquad (1)$$

Mapping of n-bits in the SPN can be defined as ***S:*** ***X->Y***, where $X = [x_1 x_2 ... x_n], x_i \in \{0,1\}$ represents input bits and $Y = [y_1 y_2 ... y_n], y_i \in \{0,1\}$ represents output bits. We have an SPN scheme, in Fig.1 with N=16, R=4 and n=4.

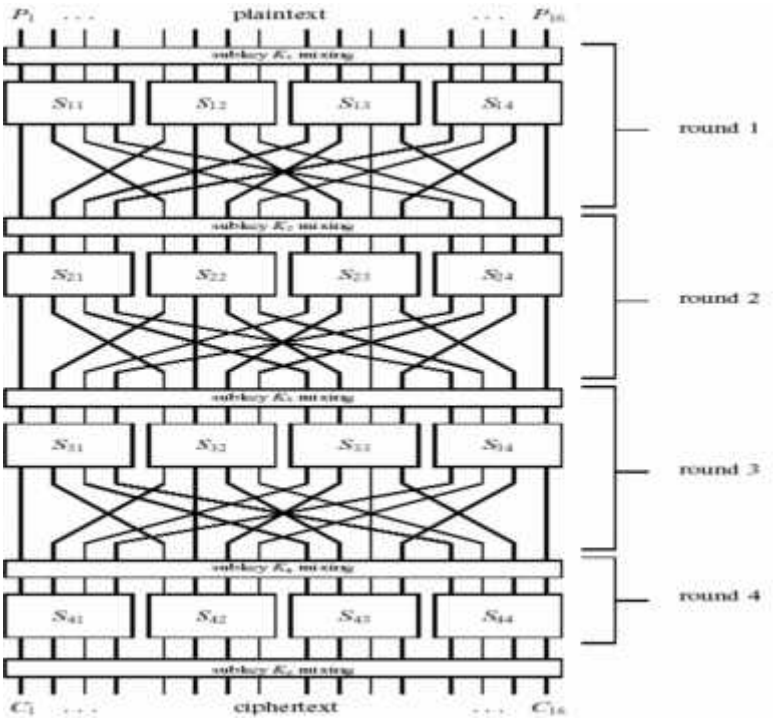

Figure 1 An SPN scheme with 4 rounds

Within a cryptosystem the s-box non-linearity is very important. That means that the function between input n-bits and output m-bits mustn't be linear. A Boolean function is like that, $f: Z_2^n \to Z_2$, in which we have only one result for every possible combination of n Boolean variables. An affine function (f) of $X=(X_{1...}X_n)$ is as in equation (2) :

$$f(X) = b_1 \otimes X_1 \oplus b_2 \otimes X_2 \oplus ... \oplus b_n \otimes X_n \oplus c = w.X \oplus c. \qquad (2)$$

The values of $b_{1,}$ $b2_{...}$ bn, c are from Z_2. If c=0 then we have a linear function f(X).

3.2 S-Boxes (Substitution Boxes)

An S-Box can be described with $Z_2^n \to Z_2^m$ mapping. This, S can be also written like that $S(X) = \left(f_1(X) ... f_m(X)\right)_{1xm}$, where the functions $f_1(X)$, $f_2(X)$ and $f_m(X)$ are Boolean ones and for that reason they can be described as $Z_2^n \to Z_2$. An S-box must satisfy some criterias to be used in algorithm. Let see these criterias:

3.2.1 Completeness Criteria

In 1979 Kam and Davida determined the completeness criteria for function $f : \{0,1\}^n \to \{0,1\}^n$. The f function is complete , for every $X \in \{0,1\}^n$ if $f(X)$ and

$f(X \oplus e_i)$ are different from each other. The avalanche vector of an S-box is given as in equation (3) below:

$$aY^{e_i} = f(X) \oplus f(X \oplus e_i) = \left[a_1^{e_i} a_2^{e_i} \ldots a_n^{e_i}\right] \quad (3)$$

The entire changing of j-th bit of the avalanche vector is calculated with equation (4):

$$W\left(a_j^{ei}\right) = \sum_{\substack{foralXeqhole \\ inputalphabet}} a_j^{e_i} \quad (4)$$

The maximum value of this is 2^n.

$$0 \le W\left(a_j^{ei}\right) \le 2^n \quad (5)$$

As a result from equation (5) an S-box to satisfy the completeness criteria must satisfy the equation (6):

$$0 < \frac{1}{2^n} W(a_j^{e_i}) < 1 \quad (6)$$

,for every value of i and j. If the value of this inequation is equal to 0 or 1 the S-box is non-complete, so it does not satisfy the completeness criteria.

3.2.2 Avalanche Criteria (AVAL)

This criteria is represented by Feistel. He pointed out that a little number of differences between bits in plaintext can cause a large number of differences between bits in ciphertext. So in a function $f : \{0,1\}^n \to \{0,1\}^n$ the changing of one input bit must cause almost changing of behalf of output bits. In conclusion an S-box to satisfy the AVAL the equation (7) must be true:

$$\frac{1}{2^n} \sum_{j=1}^{n} W(a_j^{e_i}) = \frac{n}{2} \quad (7)$$

3.2.3 Strict Avalanche Criteria (SAC)

In 1985, Webster and Tavares represented the SAC criteria merging the criterias of Completeness and AVAL. In a function $f : \{0,1\}^n \to \{0,1\}^n$ SAC criteria is true when the changing of input bits (i) will change the output bits with a propability equal to ½. So the S –box for the SAC must satisfy the equation (8):

$$\frac{1}{2^n} W(a_j^{e_i}) = \frac{1}{2} \quad (8)$$

The AVAL and SAC are very similar to each other. If an S-box does satisfy the SAC must also satisfies the AVAL, and if an S-box does satisfy the AVAL this doesn't mean that it also satisfies the SAC.

3.2.4 The Bit Independence Criteria (BIC)

This criteria is determined from Webster and Tavares. The $f : \{0,1\}^n \to \{0,1\}^n$ function for all values of $i, j, k \in \{1,2...n\}$, in case that j ≠ k. BIC criteria is about changings of the output bits (**j and k**) in a independent way from each other affected by input bits (**i**). To measure the independence concept we have to calculate the correlation coefficient between j and k bits of the avalanche vector (aY^{e_1}) with the formulation (9):

$$corr(v,w) = \frac{E\{vw\} - E(v)E(w)}{\sqrt{\left(E(v^2) - E(v)^2\right)\left(E(w^2) - E(w)^2\right)}} \quad (9)$$

E(v), is the expected value of the avalanche value v as calculated in equation (10):

$$E(v) = \frac{1}{2^n} \sum_{\substack{foralXeqhole \\ inputalphabet}} v(X)$$

(10)

The bits of an avalanche vector of an S-box can be represented as $\left[a_1^{e_i} a_2^{e_i} ... a_n^{e_i}\right]$. The changing of the i-th bit and the effect of this one through the j. and k. bits of the avalanche vector can be described in equation (11):

$$BIC(a_j, a_k) = \left| corr\left(a_j^{e_i}, a_k^{e_i}\right) \right| \quad (11)$$

for all values of 1≤*i*, j, k≤n.

3.3 Testing S-Box Properties

To test the properties of an S-box we have to analyze s-boxes step by step. Let see an example.

1. Changing one value in input bits i of an S-box, as a result we obtain the e_i bits.

2. Then we have a square matrix called the Difference Matrix, D in which the elements are obtained adding the bits of the avalanche vectors. We can calculate the values of d_{ij} using the equation (12) :

$$d_{ij} = \frac{1}{2^n} W\left(a_j^{e_i}\right) \quad (1 \le i, j \le n) \tag{12}$$

3. We have to control bits for the completeness(13), AVAL(14) and SAC(15) criterias:

- If $\exists d_{ij} = \{0,1\}$ then the s-box is not complete. (13)
- If $\sum_{j=1}^{n} d_{ij} = \frac{n}{2}, \forall i$ then the s-box does satisfy the AVAL. (for every row) (14)
- If $\sum_{j=1}^{n} d_{ij} = \frac{1}{2}, \forall i, j$ then the s-box does satisfy the SAC. (for every value) (15)

4. The correlation coefficients between j.th and k.th of Avalanche vectors are the elements of B (Bit Independence Parameters Matrix). This matrix can be defined as below:

$$B = \begin{bmatrix} b_{1.12}.....b_{1,1n} & b_{1,23}.....b_{1,2n} & ..... & b_{1,(n-1)n} \\ b_{2.12}.....b_{2,1n} & b_{2,23}.....b_{2,2n} & ..... & b_{2,(n-1)n} \\ ..... & ..... & ..... & ..... \\ b_{n.12}.....b_{n,1n} & ..... & ..... & b_{n,(n-1)n} \end{bmatrix}_{nx\binom{n}{2}} \tag{16}$$

The values of elements of the above matrix can be calculated using the equation (17):

$$\underset{1 \le i,j,k \le n}{b_{i,jk}} = BIC\left(a_j^{e_i}, a_k^{e_i}\right) \tag{17}$$

5. If the value of BIC(f) is different from 1 the s-box does satisfy the BIC criteria. If the value of BIC(f) is near to 0, the correlation coefficient between bits of the avalanche vectors will be small, and it is exactly what we want. We have an S-box 4x4.

$$aY^{e_i} = Y + f\left(x + e\right) \tag{18}$$

The values of e_1, e_2, e_3, e_4 are e_1=1000 e_2=0100 e_3=0010 e_4=0001.

Table 1 The S-box

X	Y	F(X⊕ e_1)	f(X⊕ e_2)	f(X⊕ e_3)	f(X⊕ e_4)	aY^{e_1}	aY^{e_2}	aY^{e_3}	aY^{e_4}
0000	1110	0011	0010	1101	0100	1101	1100	0011	1010
0001	0100	1010	1111	0001	1110	1110	1011	0101	1010
0010	1101	0110	1011	1110	0001	1011	0110	0011	1100
0011	0001	1100	1000	0100	1101	1101	1001	0101	1100

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

1. $W(a_1^{e_1}) = 12 \quad W(a_2^{e_1}) = 12 \quad W(a_3^{e_1}) = 12 \quad W(a_4^{e_1}) = 12$

$W(a_1^{e_2}) = 8 \quad W(a_2^{e_2}) = 8 \quad W(a_3^{e_2}) = 12 \quad W(a_4^{e_2}) = 8$

$W(a_1^{e_3}) = 4 \quad W(a_2^{e_3}) = 12 \quad W(a_3^{e_3}) = 8 \quad W(a_4^{e_3}) = 12$

$W(a_1^{e_4}) = 12 \quad W(a_2^{e_4}) = 8 \quad W(a_3^{e_4}) = 8 \quad W(a_4^{e_4}) = 8$

2. D: $d_{ij} = \frac{1}{2^n} W\left(a_j^{e_i}\right)$,j and i take the values 1,2,3 and 4. Using the equation (12) we obtain the D matrix:

$$D = \begin{bmatrix} d_{11} & . & . \\ d_{21} & . & . \\ . & . & d_{nn} \end{bmatrix}_{nxn} \longrightarrow D = \begin{bmatrix} 0{,}75 & 0{,}75 & 0{,}75 & 0{,}75 \\ 0{,}5 & 0{,}5 & 0{,}75 & 0{,}5 \\ 0{,}25 & 0{,}75 & 0{,}5 & 0{,}75 \\ 0{,}75 & 0{,}5 & 0{,}5 & 0{,}5 \end{bmatrix}$$

- The values in D are different from 1 and 0, so the s-box is complete.
- If we add for every row the values we will have 3, 2.25, 2.25 and 2.25(not equal to 2). So the S-box does not satisfy the AVAL.
- All the values of the matrix are not equal to ½ the s-box does not satisfy the SAC.

3. Then we obtain the **BIC** matrix using the equation (11) from above in text:

$$B=\begin{bmatrix} b_{1,12} & b_{1,13} & b_{1,14} & b_{1,23} & b_{1,24} & b_{1,34} \\ b_{2,12} & b_{2,13} & b_{2,14} & b_{2,23} & b_{2,24} & b_{2,34} \\ b_{3,12} & b_{3,13} & b_{3,14} & b_{3,23} & b_{3,24} & b_{3,34} \\ b_{3,12} & b_{3,13} & b_{3,14} & b_{3,23} & b_{3,24} & b_{3,34} \end{bmatrix} \rightarrow B=\begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0{,}5 & \sqrt{3}/3 & 0{,}5 & 0 & 1 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3\sqrt{3}/3 & \\ 0 & \sqrt{3}/3 & \sqrt{3}/3 & 0{,}5 & 0 & 0 \end{bmatrix}$$

3.4 Software (In Visual Basic)

In the main form are represented the input and output bits of an S-box, in Figure 2.

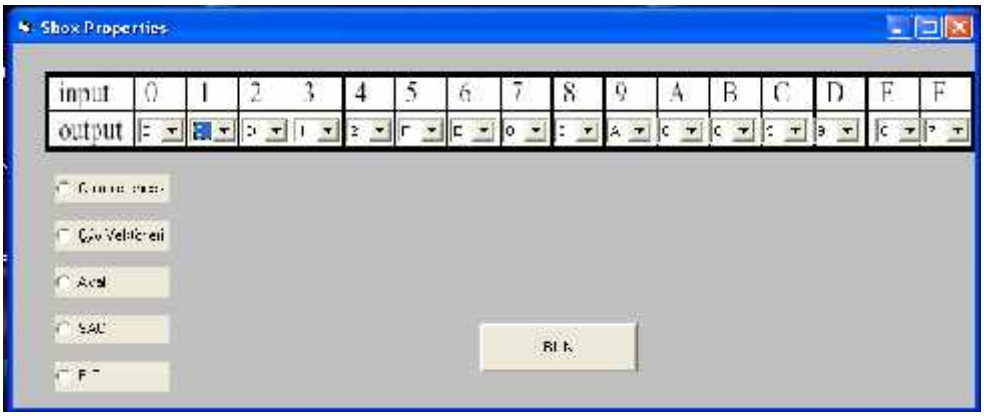

Figure 2 The main form

To determine if the S-box is or not complete we used the first option "Completeness" then the button RUN. In the same way we used the other options: Avalanche vectors, Aval, SAC and BIC and in every case we obtain another form.

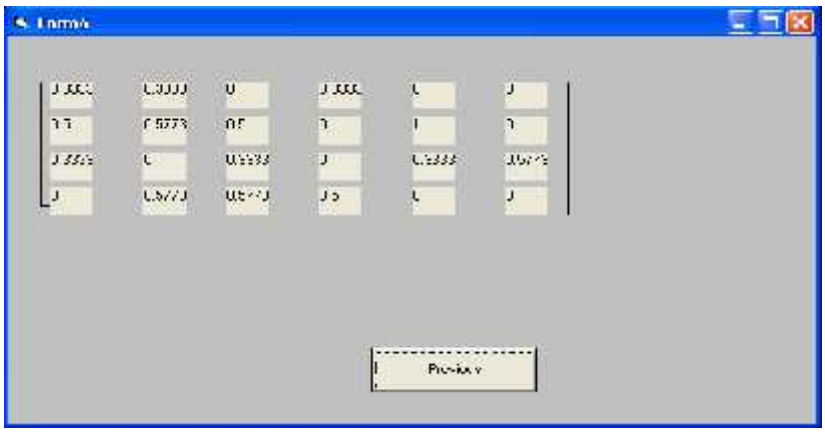

Figure 3 The form that presents the BIC matrix

Conclusion

As we see it takes time to test the properties of an S-box. In the example above the s-box doesn't satisfy the criterias and in the end (Fig.3) the values of correlation coefficient are near to 1. With the software we can change the input bits and see what we get in every special case.

References

[1] "An overview of Chryptography" Gary Kessler

[2] FIPS 197, Advanced Encryption Standard, Federal Information Processing. Standard (FIPS), Publication 197, National Bureau of Standards, U.S. Department of Commerce, Washington D.C., November 26, 2001.

[3] Bruce Schneier, Applied Cryptography: Protocols, Algorithms, and Source Code in C, U.S.A: John Wiley&Sons, 1996.

[4] W. Stallings, Cryptography and Network Security: Principles and Practises, U.S.A: Prentice Hall, 1999.

[5] N. Ferguson and B. Schneier, Practical Cryptography, U.S.A: Wiley Publishing, 2003

[6] A.F.Webster, "Plaintext/Ciphertext Bit Dependencies in Cryptographic Systems", Master's Thesis, Department of Electrical Engineering, Queen's University, CANADA, 1985.

[7] A.F.Webster and S.E.Tavares, "On the Design of S-boxes", Proc. Of CRYPTO'85, Springer-Verlag, 1985