Security Analysis of Cryptographic Algorithms by means of Boolean Functions

E.G.BARDIS*, N.G.BARDIS*, A.P.MARKOVSKI*, A.K.SPYROPOULOS**

* Department of Computer Science
National Technical University of Ukraine
(Kiev Polytechnic Institute)
Glyfada-Athens
Tainarou 66
16561
HELLAS
** Department of Mathematics
University of Athens

Abstrac t The aim of the study is to develop an approach to analysis of the security level of cryptographic algorithm on the basis of Boolean function tool application. It has been shown that security of the broad class of cryptographic algorithms, among which DES, IDEA, SHA and many others making use of bit transforms, is based on the difficulty to solve an analytically intractable problem of finding the roots of a nonlinear Boolean equation system. The break methods for cryptographic algorithms of this class are identical with strategy of search diminishing at finding the roots of such a system. Criteria for evaluation of equivalent Boolean equation system solution difficulty are worked out as well as means of their practical identification are presented. Verification of cryptoresistance of the widespread algorithms for information security such as DES and SHA has been carried out with application of he suggested approach.

Key-Words: - Cryptography, security systems, security level, Boolean Functions, nonlinear Boolean functions. *IMACS/IEEE CSCC'99 Proceedings,* Pages:3121-3126

1 Introduction

The problem of the information security has arisen with the wide development of information and communication systems. The rapid development of PCs has resulted in increase of their performance, possibility of their interconnection into networks with parallel processing at a high speed, has leaden to decrease of most frequent cryptographic algorithm security. In its turn, a great number of new algorithms appeared last years. The problem of the objective evaluation of security is considered nowadays as a basic and complicated one.

Quality of cryptographic algorithm depends on settling two contradicting demands:

- Attaining of the maximal possible security level.
- Attaining of the maximal possible speed of cryptographic data processing in PC.

The problem of the objective determination of the security is considered to be more complicated than that of speed of processing. Cryptanalysis, in practice, deals with two categories of attacks:

- "Recovery" of the plaintext by the known ciphertext and algorithm and unknown key.
- "Finding" of the unknown key, by the known plaintext, ciphertext and algorithm.

Problems of the first category can be solved only with the method of statistic analysis, while problem of the second category, which are easier, can be solved either with selective methods or with analytical ones.

In cryptanalysis problems of the second category are more frequent, since many protocols of identification and key exchange suggest transfer of both a plaintext and ciphertext through the communication channels [6].

Referring to one-way algorithms, the «break» problem may be defined as that of text recovery whose signature computed by a known algorithm coincides with the given signature code. To solve

this problem, both searching methods and those of analytic reverse transformation are used [6].

2 Problem Formulation

Determination of the objective level of cryptographic algorithm resistance to «break» by analytical and combined methods is in fact an important problem. By now this problem is estimated in most case only regarding to search. Difficulty to work out a unified strategy for estimation of cryptographic algorithm resistance to analytical "break" methods may be explained in the following way:

- Each of algorithms despite of general properties has its special features which obstacle or to the contrary favour to application of some analytical "break" methods.
- There is a wide enough scope of analytical "break" methods whose efficiency depends not only on the algorithm being broken but also on particularities of application protocols.

At present for estimation of cryptographic algorithms resistance to analytical «break» an approach is popular which consists in open publication of an algorithm and its discussion in the form of conferences issues of article collections on the subject [6]. However, such an approach is not always justified and is connected with long delay of its putting into practice as well as it can not assure that in the course of time an effective method of its «break» will not be found.

For example, DES was published at the end of 70-ties and only in the middle of 90-ties rather effective methods of its «break» were published [1,5].

That is why working out of a cryptoresistance evaluation method witch does not depend on the way of «break» is an urgent problem. This method should be founded on a general mathematic principle laid in the basis of a cryptographic algorithm. In the presented study such a method of cryptoresistance evaluation which does not depend on the «break» method is suggested for a broad class of algorithms which make use of confusion and diffusion operations. These algorithms develop the concept of Shannon [7] about "ideal cryptography". In fact all algorithm being utilized in practice, except the public key one, belong to their number. This is the reason of topicality of the given study.

3 Problem Solution

For solution of the urgent problem of development of a unified method for estimation of rather a broad class of cryptographic algorithms resistance to analytical means of «break» the propound below approach based on general enough procedures of cryptographic transformations and reducing the most part of the «break» problems to a single mathematical problem is suggested.

Analysis of a great number of cryptographic algorithms applied in modern information security systems makes it possible to conclude that essentially all of them have a block structure, so the h-bit binary vector $\overline{Y}=(y_1,y_2,\ldots,y_h)$ of the output message depends only on the n-bit binary vector $\overline{X}=(x_1,x_2,\ldots,x_n)$ of an input message and on the r-bit binary vector of the key $\overline{K}=(k_1,k_2,\ldots,k_r)$, in this case for irreversible algorithms n=h. Irrespective of the cryptographic algorithms being used, the components $y_i \in \{0,1\}$, i=1,…,h of the output vector $\overline{Y}$ may be represented by the system of h Boolean functions.

$$\begin{aligned} &f_1(x_1,x_2,\ldots,x_n,k_1,k_2,\ldots,k_r)=y_1\\ &f_2(x_1,x_2,\ldots,x_n,k_1,k_2,\ldots,k_r)=y_2\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &f_h(x_1,x_2,\ldots,x_n,k_1,k_2,\ldots,k_r)=y_h \end{aligned} \qquad (1)$$

In action, each function of system (1) is complicated enough and for the most part of cryptographic algorithms it comprises thousands of terms. For some algorithms which do not operate with the key (for example for SHA) the functions of system (1) depend only on the computer vector $\overline{X}$. If at symmetrical algorithm cryptanalysis values of the input informational vector $\overline{X}$ and of the key vector $\overline{K}$ are cryptographic transformation unknown elements and there are no practical restrictions on the number of known output blocks, then system (1) of Boolean functions is transformed into the system of h Boolean equations with n+k unknowns:

$$\begin{aligned} &p_1(x_1,x_2,\ldots,x_n,k_1,k_2,\ldots,k_r)=0\\ &p_2(x_1,x_2,\ldots,x_n,k_1,k_2,\ldots,k_r)=0\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &p_h(x_1,x_2,\ldots,x_n,k_1,k_2,\ldots,k_r)=0 \end{aligned} \qquad (2)$$

At solution of the second «break» problem, when values of the vectors $\overline{X}$ and $\overline{Y}$ components are known, system (1) of Boolean functions transforms into the following system of Boolean equations:

$$\begin{aligned} &\varphi_1(k_1,k_2,\ldots,k_r)=0\\ &\varphi_2(k_1,k_2,\ldots,k_r)=0\\ &\ldots\ldots\ldots\ldots\ldots\\ &\varphi_h(k_1,k_2,\ldots,k_r)=0 \end{aligned} \qquad (3)$$

The "break" problem is equivalent to solution of presented system (3) of Boolean equations with respect to $k_1,k_2,\ldots,k_r$.

As applied to one-way algorithms, when the key $\overline{K}$ (if it is being utilized) and the output vector $\overline{Y}$ are the known elements, system (1) of Boolean function is reduced to the system of Boolean equations:

$$\begin{aligned}&\psi_1(x_1,x_2,\ldots,x_n)=0\\&\psi_1(x_1,x_2,\ldots,x_n)=0\\&\ldots\ldots\ldots\ldots\ldots\\&\psi_h(x_1,x_2,\ldots,x_n)=0\end{aligned} \qquad (4)$$

In this case as well the "break" problem is equivalent in the mathematical aspect to solution of system (4) of Boolean equations.

Thus, the problem of symmetrical reversible and one-way cryptographic algorithms "break" is reduced to solution of a Boolean equation system.

Solution of a linear Boolean equation system does not involve great difficulties. In particular, Madryga [4] symmetrical reversible cryptographic transform algorithm does not utilize nonlinear functional transforms and, consequently, equivalent to it system (1) comprises only linear functions. At "break" of this cryptographic algorithm, when unknown elements are n+r components of the $\overline{X}$ and $\overline{K}$ vectors, the problem is reduced to solution of system (2): as the number of this system equations is less by r than the number of unknowns, search of 2^r versions must be carried out for its practical solution. At solution of the second "break" problem, the corresponding system (3) of Boolean equations may be solved analytically with respect to the key vector components. In such a way the indicated cryptographic algorithm by Madryga may be easily broken if a portion of an input message is known.

However, solution of nonlinear Boolean equation system is known to be a complicated mathematical problem which in full extend may be referred to the category of Boolean algebra problems difficult to solve. The only way of their solution is searching. Under certain conditions, with application of some special analytical methods the search area may be substantially decreased. Conditions of mentioned methods efficiency are fully determined by the form of Boolean functions.

Let us consider in more details solution of nonlinear Boolean equation systems. The most widespread in practice method is that of linear approximation. Its idea consists in the fact that each of the nonlinear Boolean functions composing system (1) is substituted for the nearest by the Hamming distance linear function. In such a way transition from system (1) of nonlinear Boolean equations to the system of linear ones is carried out:

$$\begin{aligned}&\phi_1(x_1,\ldots,x_n)=y_1\\&\phi_2(x_1,\ldots,x_n)=y_2\\&\ldots\ldots\ldots\ldots\ldots\\&\phi_h(x_1,\ldots,x_n)=y_h\end{aligned} \qquad (5)$$

with $\phi_j(x_1,\ldots,x_n)=a_{0j}\oplus a_{1j}\cdot x_1\oplus a_{2j}\cdot x_2\oplus\ldots\oplus a_{nj}\cdot x_n$. The function is chosen starting from the following condition:

$$\phi_j(x_1,\ldots,x_n)=\min_{j,\ x\in Z}\sum\phi_j(x_1,\ldots,x_n)\oplus f_j(x_1,\ldots,x_n) \qquad (6)$$

System (6) of linear Boolean equations may be solved analytically and the vector $\overline{X}_o=\{x_{01},\ldots,x_{0n}\}$ may be obtained. The true solution of the equation system (1) should be sought by search, starting from the vector $\overline{X}o$ and increasing step-by-step in the course of search the Hamming distance from the vector of reference. The Hamming distance between the solution vector of systems (1) and (5) depends on the total non-linearity of all the Boolean functions entering into system (1), therefore, for search increase the total non-linearity of all the Boolean functions should be increased to the maximal extend and for this purpose non-linearity of each Boolean function must be maximal. If a cryptographic algorithm is not resistant enough to the analytical "break" methods, then the equations of the Boolean function system equivalent to it have a low non-linearity and may be replaced by their linear approximations, that in the end will make it possible to decrease the search by many times and to solve the "break" problem for a reasonable time. In particular, the widely known DES - algorithm has been broken [5] by the methods of combined application of linear approximation and directed search with an increasing Hamming distance.

Another approach to information security algorithm "break" often used in the practice of cryptanalysis is the method of variables excluding. The idea of the method consists in analysis of significance of each of the variables in the equations and those with the lowest significance are excluded. In this case the significance is usually estimated statistically or by applying the probability theory approach. After the variables of low significance have been excluded, decreasing the dimensionality of Boolean equation system solution problem is

attained and further on the system is solved by the search methods.

To decrease efficiency of the described method for equation system solution it is necessary that each of the variables be maximal significant, or that is the same, after excluding any variables from the function $f(x_1,\ldots,x_n)$ the function must be balanced. This demand may be referred to as the condition of the maximum of conditional entropy of each Boolean function entering into system (1) On account of great importance of the indicated criterion for theory and practice of cryptographic information security systems, it has been specially named as Strict Avalanche Criterion or SAC for short.

Thus, to intricate to the maximal extend the solution of Boolean equation equivalent systems by means of variables excluding, each Boolean function of the system must correspond to SAC. In cryptanalysis practice exclusion of variables and diminishing the dimensionality of a problem is most often attained by combination with other methods, in particular, in the form of wide known method of differential cryptanalysis [3].

The method of differential cryptanalysis itself presents conceptually a way of non-linear Boolean equation system solution, based on probablistic "restoring" of Boolean functions composing a system. Differential cryptanalysis can not be applied directly to solution of systems (2) – (4) of nonlinear Boolean equations as to a mathematical problem, i.e. when the number of Boolean equations is restricted to h. In cryptanalysis action, as a rule, the party carrying out cryptanalysis has at its disposal rather a great amount of output message blocks, so the equation number m surpasses h considerably. It is this circumstance that makes it possible in the framework of differential analysis to fulfil probablistic "restoring" of a Boolean function at which probablistic determination of unknown variables is being carried out. Furthermore, if the Boolean functions composing a system are "poor" in the cryptographic aspect, then the "restoring" method demands a less number m of specific equations comparing to the ordinary search method. A poor function is called a such one which on excluding of a variable turns to be a partial Boolean function without the maximum of entropy, i.e. which takes values of zero and one with non-equal probability. This in fact makes it possible to determine whether the variables are equal to 0 or 1 with probability different from 0.5. Thus, efficiency of differential cryptanalysis is determined by thy type of the Boolean functions composing system (1): if these functions possess the maximum of total and conditional entropy, then computational expenditures approach the total search.

Another way of nonlinear Boolean equation system solution is based on diminishing of search at the cost of statistic analysis of the cipher block output vector. To ensure the minimal productivity of this "break" method, it is necessary that each of output code bits assume values of zero or one with equal probability at any input code. In other words, to protect a Boolean equation system from statistic methods of solution, each of the Boolean functions of the system equivalent to the cryptographic algorithm must have the maximum of total entropy, or that is the same, must be balanced (must take the value of one on a half of the possible variable sets).

Solution of Boolean equation systems may be significantly simplified if the Boolean functions which make up system (1) are correlated.

Thus, carried out analysis of nonlinear Boolean equation solution methods reveals that:

- solution can in no way be performed completely by the analytical approach and it always suppose search utilization: in this aspects all the methods considered are in fact combined, furthermore, their analytical constituent is intended to decreasing the search version number;
- all analytical methods of cryptographic algorithm "break" are, in essence, taking into consideration the cryptanalysis specific conditions, nonlinear Boolean equation system solution strategies;
- difficulty of nonlinear Boolean equation system solution by the combined method outlined above is determined completely by specific properties of the Boolean functions making up a system.

The principal difference of "break" problems from the mathematical task of Boolean equation system solution consists in the fact that the number of equations is much more than the number of variables.

Obviously, that solution of system (2) is much more intricate comparing to that of system (3) which are equivalent respectively to the first and second problems of cryptographic algorithm "break". However, in practice, as it has been noticed above, the second problem of cryptanalysis is more often met. Its partial version is the problem for which it is possible to preset arbitrary an input informational message portion. In this case there appears an opportunity to chose such $x_1,x_2,\ldots,x_n$ that, at their substitution into equations of system (1), system (3) of Boolean equations with the minimal intricacy of its solution must be obtained. That is why another important criterion of cryptographic algorithm

"break" resistance consists in understanding that at any set of Boolean variables $x_1,x_2,\ldots,x_n$ equation system (1) must transform into equation system (3) or (4) which should correspond to the demand of maximal difficulty of their solution.

Thus, the conclusion seems to be substantiated that estimation of cryptographic algorithm "break" difficulty by means of analytical and combined methods may be carried out through estimation of difficulty to solve the Boolean function system which is equivalent to the algorithm. This difficulty, in its turn, may be determined through specific properties inherent in the Boolean functions composing the system equivalent to the algorithm. If these specific properties are inherent in full measure in each Boolean function making up system (1), then solution of such a system is intricate to the extreme extent in computational aspect: the search decrease methods considered above appear to be ineffective and time consumption at their application does not differ in essence from that at total search.

The analysis of the Boolean equation system solution methods presented above reveals that maximal difficulty of solution is attained if the following conditions are held:

1. Each of the Boolean functions $f_i(x_1,\ldots,x_n, k_1,\ldots,k_r)$, $\forall$ i=1,…h of the system equivalent to the cryptographic algorithm of system (1) must have property of maximal total entropy, that is be balanced.
2. Each of the Boolean functions $f_i(x_1,\ldots,x_n, k_1,\ldots,k_r)$, $\forall$i =1,…h of the system equivalent to the cryptographic algorithm of system (1) must have the maximum of conditional entropy, that is the partial function obtained at excluding of any variable $x_1,\ldots,x_n,k_1,\ldots,k_r$ must have the maximum of entropy or, that is the same, be balanced.
3. Each of the Boolean functions $f_i(x_1,\ldots,x_n, k_1,\ldots,k_r)$, $\forall$i =1,…h of the system equivalent to the cryptographic algorithm of system (1) must have the maximal value of non-linearity, that is be a bent-function.
4. Boolean functions composing system (1) equivalent to the cryptographic algorithm must be non-correlated in pairs or, that is the same, the Boolean function
   $\zeta_{ij}(a,x_1,\ldots,x_n,k_1,\ldots,k_r)=a\cdot f_i(x_1,\ldots,x_n,$
   $,k_1,..,k_r)\oplus a\cdot f_j(x_1,\ldots,x_n,k_1,\ldots,k_r)\oplus$
   $\oplus f_j(x_1,\ldots,x_n,k_1,\ldots,k_r)$, $\forall$j,i=1,…,h, i≠j must have the maximum of conditional entropy with respect to all the variables on which it is determined.
5. The number h of the system equations must be as great as possible, that is the capacity of the blocks being processed with a cryptographic algorithm must be as great as possible.

It should be especially pointed out that the suggested approach is applicable to the full extent to estimation of "break" resistance of rather a broad class of cryptographic algorithms, that means it includes both symmetrical reversable algorithms and algorithms of hash-signature formation based on the operations of confusion and diffusion [7].

It may be said from the theoretical aspect that the cryptographic foundation of such algorithms is the analytically intractable problem of finding the roots of nonlinear Boolean equation system.

The suggested approach to estimation of the security level can not be applied to cryptographic algorithms in the basis of which other analytically unsolvable mathematical problems lie. This is true for public-key algorithms such as RSA, EL-Gamal and ECC.

In essence two cryptographic algorithms, namely DES and SHA have been investigated by the method suggested.

The investigation of DES has revealed that equivalent Boolean functions are balanced, non-correlated in pairs, correspond to SAC, however their non-linearity is less than maximal possible. It follows here from that the most effective method of DES "break" is linear cryptanalysis [5], which demand in average 2^{43} tries while the total search does 2^{56} tries. Investigation of SHA has shown that the equivalent Boolean functions also satisfy all the cryptographic criteria except that of maximal non-linearity. Analysis has reavealed that at this cost search may be decreased by 3 - 4 orders, however great capacity (160) of hash signatures makes this method of linear approximation impossible for im-plementation at present time.

4 Conclusion

Growth of PC performance makes it possible to solve in practice the problem of analysis of cryptographic algorithms on the microlevel, i.e. on the level of Boolean functions. Any algorithm based on operations of confusion and diffusion can be presented as a system of Boolean functions by formalized methods. Solution of the problem of cryptanalysis tends to solution of a system of Boolean equations. Finding the roots of nonlinear Boolean equation systems is referred to a number of mathematical problems which can not be solved with analytical methods. The only method for their solution is search. Thus it has been shown that the

foundation of cryptographic properties of algorithms based on operations of confusion and diffusion is the analytically intractable problem of finding the nonlinear Boolean equations roots. The search area at solution of this problem can be reduced by special means making use of properties of the Boolean functions composing the system. Connection between these methods and those of cryptographic algorithm “break” is shown in the study. In certain conditions which the Boolean functions must hold, the search area may be considerably diminished. If the Boolean functions of the system equivalent to the cryptographic algorithm satisfy this condition, then all the known cryptanalysis methods appear to be ineffective and in practice result in total search. Therefore, cryptoresistance of algorithms is suggested to be estimated through analysis of Boolean function systems formed from bit transformations.

The advantage of this approach consists in the fact that it can estimate cryptoresistance of a broad class of algorithms with application not only the total search but also other “break” methods.

Practical applicability of suggested approach is proved by experimental study of DES and SHA algorithms security.

References:

[1]D.Coppersmith. The Data Encryption Standart (DES) and its Strength against attacks. IBM Jornal of Research and Development, Vol.38, No.3, 1994, pp.243-246

[2]J.Desmedt, J.Quisquater, and M.Davio, Dependence of Output on Inouts in DES: Small Avalanche Characteristics, Advances in Cryptology-CRYPTO ’84 Proceedings, Berlin:Springer-Verlag 1985,pp. 359-376.

[3]E.Biham and A.Shamir, Differential Cryptanalysis of FEAL and N-Hash, Extended Abstract, Proceedings of EUROCRYPT’ 91, 1991, pp.102-125.

[4]W.E.Madryga. A high performance encryption algorithm. Computer Security: A Global Challenge, Elselvier Science Publishers. 1984. pp.557-570.

[5]M.Matsui, Linear Cryptanalysis Method for DES Cipher, Proceedings of EUROCRYPT ’93, Spriger-Varlag,1994, Vol.765, pages 386-397

[6]B.Schneier. Applied Cryptography. Protocols. Algorithms and Source codes in C. Ed.John Wiley, 1996, pp.758.

[7]C.E.Shannon. Communication theory of secrecy systems. Bell Technical Journal, Vol.28, No.10, 1949, pp.656-715.