International Journal of Cryptology Research 1 (1): 65- 72 (2009)

Random Walk Algorithm Based Design Technique for S-Box

S. Kazmi and N. Ikram
National University of Science and Technology, Pakistan
Email: shaguftakazmi_2007@hotmail.com

ABSTRACT

Substitution boxes (S-Boxes) are important in the design of symmetric algorithms. Design of efficient and cryptographically strong S-boxes have always been an area of active research. In this paper, we introduce a new cryptographic method for generating substitution boxes based on Graph Theory, which is efficient and has been analyzed to possess good cryptographic properties such as Differential Probability, Linear Probability, Algebraic Degree and particularly higher "Best Linear Approximation". The algorithm is based on rapidly mixing random walks. We show that an S-Box that meets the criterion of random walk is invulnerable to the cryptanalysis because of satisfying various cryptographic properties along with randomness.

INTRODUCTION

S-Boxes are important in the design of secure symmetric-key algorithms. These are designed prudently with high resistance against many statistical and cryptographic attacks as well as with low implementation cost on several platforms. In this paper, we present a new method on the design of S-Box using the basic concepts of graph theory. The study of random graphs started with the influential work of Erdos and Renyi in the 1950s and 1960s (Bollobas, 1985). Random graph theory has turned into a prop of modern Discrete Mathematics. With particular focus on cryptographic and statistical properties, we present a design criterion for cryptanalysis-resistant S-Box using scaled up Random Walk based Algorithm (RWA). The design analysis characteristically involves analysis of simple statistical parameters such as path length, mixing time, degrees distribution and graph distribution. The random walk over its finite state space includes different steps iterated in the main set of commands of the RWA. This can be modeled by a graph G with $N(=64)$ nodes, and edges connecting nodes that differ by a transposition. The sections that follow in this paper give the random walk based algorithm, design rationale including operations used in algorithm, rapidly mixing random walks, degree distribution and concentration as subsections. The summary and results are given in the last section of this paper.

RANDOM WALK BASED ALGORITHM (RWA)

In this section, we describe the RWA and its basic constituent components. The prescribed method uses a small-sized (4x4) S-Box possessing balance, Minimal Differential Probability (MDP) and Minimal Linear Probability (MLP) to generate a large-sized (8 x 8) S-Box which is generated efficiently, utilizes less memory and has improved "Best Linear Approximation (BLA)" with acceptable "Differential Probability (DP)" and "Linear Probability (LP)".

Design of S-box with good BLAs has gained importance as S-Box of Advanced Encryption Standard (AES) does not possess good BLAs for two Boolean functions i.e. $f_5 = 1 + X_2 \; and \; f_6 = 1 + X_1$. In our RWA based S-Box design, it is well improved for each Boolean function $fi(0 \le i \le 7)$. It also satisfies reasonably good DP and LP.

RWA is presented as follows; "$>>> a$" denotes the right cyclic shift by "a" bits, addition is modulo 2^4, "S" is a randomly chosen 4x4 S-Box with $MDP = 2^{-2}, MLP = 2^{-2}$ and Balance containing values such as $\{S\} = \{1,4,c,f,8,e,2,7,a,3,d,0,b,5,6,9\}$. The structure has been designed to generate good and efficient S-Box which is later analyzed to possess good cryptographic properties. This design is memory efficient, making it suitable for implementation on smart cards and similar tokens.

Input: 8-bit input state X

Output: 8-bit output state Y

for $i = 0$ to 255

$X = i$

$X[0] \leftarrow (X >> 4) \,\&\, 0xf$

$X[1] \leftarrow X \,\&\, 0xf$

for $j = 0$ to 1

$Z[j] \leftarrow X[j] >>> 3$

$O[j] \leftarrow S(Z[j])$

$D \leftarrow (Z[0] + Z[1]) \% 2^4$

$O[2] \leftarrow S(D)$

for $k = 2$ to 3

$Z[k] \leftarrow O[k-2] >>> 1$

for $m = 0$ to 1

$$D[m] \leftarrow \left(Z[m+2] + \mathrm{O}[2]\right) \% 2^4$$
$$Y[0] \leftarrow S\left(D[0]\right)$$
$$Y[1] \leftarrow S\left(D[1]\right)$$
$$Y \leftarrow Y[0] \| Y[1]$$

The use of random walks or, more precisely, pseudo-random walks is presented in figure 1. This effectively comprises of two layers each with identical steps of Rotation, Addition and substitutions.

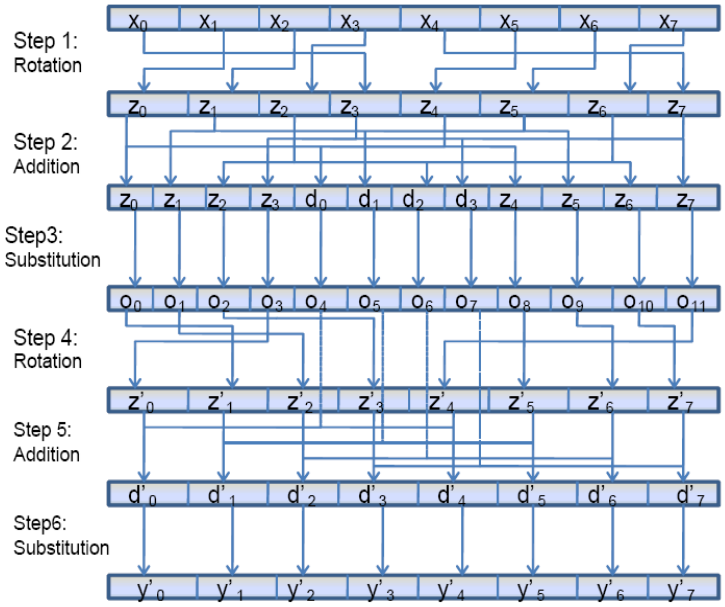

Figure 1: Path from initial state to the final state

DESIGN RATIONALE

It is important to indicate that the simple operations are extremely desirable for fast implementation. The updates of the states are based on several simultaneously applied random walks which are very simple and efficiently implementable.

Operations used in RWA

Note that the random walks are interleaved, and the randomness of each one of them relies on the randomness of the others. Also note that the updates use modular addition and not a bitwise XOR operation. This partially resolves the problem of high-probability short correlations in random walks. In an undirected random walk, there is a high probability that after a short number of steps, the state returns to a previous state, while in a directed random walk this phenomenon does not exist. The usage of addition, which unlike XOR is not an involution, prevents this property (Keller et al , 2007). Another operation used in the algorithm is bit wise rotation. We know that it is sufficient to know the k $LSBs$ of the input to retrieve the k $LSBs$ of the output. An attacker can use this property to mount an attack based on analyzing only the k $LSBs$ of the output, and disregard all the other bits. This makes the guess-and-determine attack possible with very low time complexity. For example, if no rotations were used in the algorithm, then a variant of the standard guess-and-determine attack would apply by examining only the $LSBs$ of every output word, and reducing the time complexity of the attack to the fourth root of the original time complexity (Keller et al , 2007). The third operation, substitution, provides confusion in the resultant output.

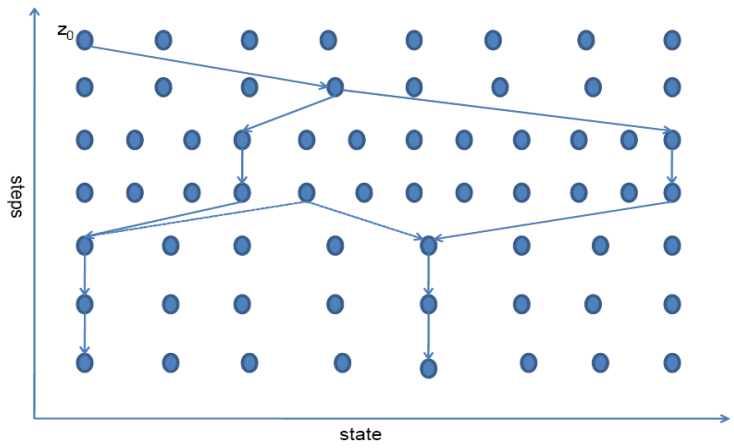

Figure 2: Subgraph showing path of a single input bit

Rapidly Mixing Random Walks

The good long term randomness properties of the internal state of algorithm are achieved by updates using rapidly mixing random walks. Since the algorithm is completely deterministic, the walks are only pseudo-random, so

we try to make the update principle like a random walk. A random walk on a graph starts at a node z_0 (see Fig2), and at each step moves to a node connected by one of its adjacent edges at random. Figure 2 shows the path from a single bit of input state to the output state bits. A random walk is called "rapidly mixing" if, after a relatively short time, the distribution of the state of the walk is close to the uniform distribution, regardless of the initial distribution of the walk (Keller et al , 2007). One of the most important parameters of RWA is its mixing time. This includes the number of steps required for a sequence of independent random moves from an initial state to an ending state achieving uniform distribution over the state space. The parameter "mixing time" is typically not easy to determine but it can be shown that the random moves achieve uniform distribution over the state space. Let's assume that our null hypothesis is H_0: The process has a specified distribution i.e. uniform distribution. We applied Pearson's Chi-Square test for goodness of fit with 7 classes (steps) and known probabilities p_i for each class that are $1/8, 1/8, 1/6, 1/4, 1/4, 1/4$ and $1/4$. The calculated value of test-statistic (x^2) is 13.17 with degree of freedom equals $6(d.f = 7-1-m)$ where m is the number of parameters estimated from the sample). In our case $m = 0$, as we have not estimated the parameters of the specified distribution. A small computed value of (x^2) indicates a good fit and it leads to the acceptance of the null hypothesis. As our calculated value is smaller than the critical value (15.09) at 1% level of significance, so we can conclude that the distribution of the state of the walk is uniform.

Degree Distribution

It is assumed that the presence or absence of an edge between two vertices is independent of the presence or absence of any other edge (see Fig2), so that each edge may be considered to be present with independent probability p. If there are $n(=13)$ vertices in a sub graph, and each is connected to an average of l edges, then it is trivial to show that $p = 1/(n-1)$, which for large n is usually approximated by $l/n\,(As\ n-1 \cong n \text{ for } n \to \infty)$ (Newman et al, 2001). The degree k for any particular vertex has a probability distribution p_k given by $p_k = {}^nC_k\,(p^k)(1-p)^{n-k}$. These probabilities are $0.381, 0.069, 0.207, 0.207, 0.207, 0.207, 0.207, 0.069, 0.069, 0.207, 0.207, 0.381$ and 0.381. For large n, the probabilities are calculated using the density function of Poisson distribution taking k as a random variable with mean l $(p_k = l^k e^{-l}/k!$ if $n \to \infty)$.

The degrees k_j of all j (for $j=1$ to n) vertices are independently identically distributed $(i.i.d)$ random integers drawn from a specified distribution with $P(k_j \le x) = F(x)$, as the graph is assumed to be entirely random. For a given choice of these degrees, also known as ''degree sequence,'' a graph is chosen uniformly at random from the set of all graphs with that degree sequence (see sec 3.2). The probability mass function and the distribution function of the vertex degree k are denoted by $P(k=j) = f_j$ for $j = 0,1,2,...,n-1$ and $F(x) = \sum_{j=0}^{|x|} f_j$ where $|x|$ is the largest integer smaller than or equal to x. Note that $f_j = (1/jc_p)\, p(1-p)^{j-2}, \forall j \ge 1$ where c_p is the normalizing constant (Hofstad, 2004). For some $t > 3$ and some positive constant c, if $1 - F(x) \le cx^{t+1}$ for $x > 0$ then it concludes that the second moment of k is finite. We denote $\mu = E[k], v = E\left[k_j(k_j - 1)\right] / E\left[k_j\right]$ and the distance or hopcount H_n between the starting and ending vertices as the minimum number of edges, is minimum. In some cases, the random graphs with appropriate distributions of vertex degree show a measurable discrepancy between theory and reality, perhaps due to the existence of additional structure in the network that is not captured by the random graph.

Thresholds: Concentration

There are a number of methods of showing concentration of a random variable that are now commonly used in random graph theory, such as Chernoff's bound, Azuma-Hoeffding inequalities, Martingale-based inequalities and Talagrand's inequalities (Hoeffding, 1963). But it is trivial to use the simplest one that is given as follows:

Chebyshev's inequality: Let X be a random variable with variance δ^2, and let $\varepsilon \ge 1$. Then

$$P\left(\left|X - E[X]\right| > \varepsilon\delta\right) < 1/\varepsilon^2$$

Using the measures of central tendency and dispersion mentioned above, we can analyze this inequality for showing concentration of the random variable k_j. For simplicity, let's assume $\varepsilon = 1$, if $P\left(\left|k_j - E[k]\right|\right) > \sqrt{v}) < 1$, then it can be shown that the particular vertex degree is concentrated close to its expected value. The probability that the random walk returns to a previously visited vertex is very important for cryptographic purposes, since short cycles lead to mathematical or statistical relations which an attacker can exploit.

RESULTS AND SUMMARY

This paper proposes a new fast and secure algorithm for generating S-Box. The main ascribed features of the algorithm are simplicity, efficiency, less memory usage in software and low implementation cost. The algorithm is based upon trivially analyzable and simple components. The algorithm possesses reasonable security in terms of the operations used. In addition, the generated S-Box possesses good cryptographic properties such as MDP, MLP and higher BLA. The algorithm involves new rapidly mixing random walks, to ensure the random walks in the long run. We also analyzed the algorithm with the help of statistical tools such as parametric distributions which can be used to find the concentration and randomness of the degrees and paths evolved through the main loop of algorithm. We have analyzed both theoretically and empirically, that the S-Box generated by the given design method has better cryptographic and statistical properties. The results are summarized in the Table 1:

TABLE 1 : Results for the designed 8 x 8 S-Box

Best Linear Approximation	**BLAP**	**Algebraic Degree**	
$f_0 = X_4 + X_3 + X_1 + X_0 + 1$	0.6094	7	
$f_1 = X_6 + X_4 + X_3 + X_2 + X_0 + 1$	0.5977	8	
$f_2 = X_7 + X_6 + X_4 + X_3 + X_1 + X_0 + 1$	0.6055	8	
$f_3 = X_7 + X_6 + X_5 + X_4 + X_1 + X_0 + 1$	0.6172	7	
$f_4 = X_7 + X_5 + X_4 + X_0 + 1$	0.6172	7	
$f_5 = X_7 + X_6 + X_4 + X_2 + X_0 + 1$	0.5977	8	
$f_6 = X_7 + X_5 + X_4 + X_3 + X_2 + X_0 + 1$	0.6133	8	
$f_7 = X_5 + X_4 + X_3 + X_2 + X_1 + X_0 + 1$	0.6250	7	
Values of Cryptographic Properties			
MDP	$2^{-4.19}$	**MLP**	$2^{-3.50}$

Here, $fi(0 \le i \le 7)$ are Boolean functions contained in the designed S-Box. Some statistical tests are applied to check the randomness of the output generated by AES algorithm by replacing its original S-box with our designed S-Box. The results show better randomness and uniformity.

REFERENCES

Bollobas, B. 1985. Random graphs, *Academic Press, London*.

Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* No. 58.

Hofstad, R.V.D, Hooghiemstra, G and Mieghemy, P.W. 2005. Distances in random graphs with finite variance degrees, *Random Structures & Algorithms*, Volume 27, Issue 1, ISSN: 1042-9832.

Keller, N; Miller, S.D; Mironov, I and Venkatesan,R. 2007. MV3: A new stream cipher based on random walks and revolving buffers, *Topics in Cryptology – Proceedings of CT-RSA 2007, Springer Verlag Lecture Note in Computer Science*, No. 4377.

Newman, M. E. J; Strogatz, S. H and Watts, D. J. 2001. Random graphs with arbitrary degree distributions and their applications. *American Physical Society Journal PHYSICAL REVIEW E,* Volume 64, 026118