IJST (2014) 38A3 (Special issue-Mathematics): 349-353
Iranian Journal of Science & Technology

http://ijsts.shirazu.ac.ir

Statistical properties of the square map

S. M. Dehnavi[1]*, A. Mahmoodi Rishakani[2], M. R. Mirzaee Shamsabad[3] and E. Pasha[1]

[1]*Faculty of Mathematical and Computer Sciences, Kharazmi University,Tehran, Islamic Republic of Iran*
[2]*Faculty of Sciences, Shahid Rajaee Teacher Training University, Tehran, Islamic Republic of Iran*
[3]*Faculty of Mathematics and Computer Science, Shahid Bahonar University, Kerman, Islamic Republic of Iran*
E-mail: std_dehnavism@khu.ac.ir

Abstract

The square map is one of the functions used in cryptography. For instance, the square map is used in Rabin encryption scheme, block cipher RC6 and stream cipher Rabbit, in different forms. In this paper, we study statistical properties of the output of the square map as a vectorial Boolean function. We obtain the joint probability distribution of arbitrary number of the upper and the lower bits of the output of square map along with the asymptotic probability distribution of the upper bits of its output. Based upon a measure for evaluating the imbalance of maps, we study the imbalance of limit distribution of the restriction of square map to its upper bits. Last, we introduce the square root map and examine this map as a vectorial Boolean function; we compute probability distribution of the component Boolean functions of this new map and also obtain the imbalance of the square root map.

***Keywords**:* Square Map; square root map; vectorial boolean function; component boolean function; asymptotic probability distribution

1. Introduction

The square map is one of the functions used in cryptography. For instance, the square map is used in Rabin encryption scheme (Stinson, Chap. 5, 2003). In this public key encryption system, the square map is computed modulo the product of two large primes. The square map is also used in block cipher RC6 (Rivest, et al. 1998). In this symmetric cipher, a quadratic polynomial over the ring $Z_{2^{32}}$ is computed. As another example, in the design of the stream cipher Rabbit (Boesgaard, et al. 2003), the square map is used. In this cipher, the square map is not modular; the square map is computed as a function over natural numbers. In fact, the square map is considered as a vectorial Boolean function from 32-bit natural numbers to 64-bit natural numbers and then the upper and the lower segments of the output of this map are XORed.

In this paper, we investigate statistical properties of the output of the (non-modular) square map as a vectorial Boolean function. We obtain the joint probability distribution of arbitrary number of the upper and the lower bits of the output of the square map along with the asymptotic probability distribution of the upper bits of its output. Then, the probability distribution of the component Boolean functions of the output of this map are obtained. After introducing a measure for evaluating the imbalance of maps, we examine the imbalance of limit distribution of the restriction of square map to its upper bits. Last, we introduce the square root map and examine this map as a vectorial Boolean function; we compute probability distribution of the component Boolean functions of this new map and also we obtain the imbalance of the square root map.

In Section 2 preliminary notations and definitions are presented. Section 3 is devoted to computing the probability distribution of the output of square map and Section 4 studies the imbalance of the square and the square root maps.

*Corresponding author
Received: 26 February 2014 / Accepted: 20 July 2014

2. Preliminary Definitions and Notations

In this paper, the number of elements or cardinality of a finite set A is denoted by $|A|$. For a function $f: A \to B$, the preimage of an element $b \in B$ is denoted by $f^{-1}(b)$ and is defined as $\{a \in A | f(a) = b\}$.

Let F_2 be the finite field with two elements. Each element of F_2^n (The Cartesian product of n copies of F_2) can be considered as a vector of length n. Each function $f: F_2^n \to F_2$ is called a Boolean function and each function $f: F_2^n \to F_2^m$ with $m > 1$

is called a vectorial Boolean function or a Boolean map; such a function can be viewed as a vector $(f_{m-1}, \dots, f_0)$ of f_i's, $0 \le i < m$. Here, f_i's are Boolean functions from F_2^n to F_2. These Boolean functions are called component Boolean functions of the vectorial Boolean function f. Also, if $x \in F_2^n$, then the i-th bit of x is denoted by x_i.

We denote the vector $(0, \dots, 0)$ by $\mathbf{0}$. For every natural number t and each function $f: F_2^n \to F_2^m$ with $m > t$, we denote the restriction of f to indices $i_0, i_1, \dots, i_{t-1}$ of the output by $(f_{i_{t-1}}, \dots, f_{i_0})$.

There is a one-to-one correspondence between Z_{2^n}, the ring of integers modulo 2^n, and F_2^n as

$$\varphi: F_2^n \to Z_{2^n}$$
$$x = (x_{n-1}, \dots, x_0) \mapsto \varphi(x) = \sum_{i=0}^{n-1} x_i 2^i;$$

this natural correspondence is used throughout this paper.

Suppose that m, n and d are natural numbers with $n = dm$. A function $f: A \to B$ with $|A| = n$ and $|B| = m$ is called balanced if and only if for every $b \in B$ we have
$|f^{-1}(b)| = d$.

Suppose that X is a random variable which is uniformly distributed on F_2^n. Every Boolean map $f: F_2^n \to F_2^m$ determines a random variable Y on the codomain of f, i.e. F_2^m; in other words, we have the induced random variable $Y = f(X)$ with

$$P(Y = y) = \frac{|f^{-1}(y)|}{2^n}, \qquad y \in F_2^m.$$

3. Computing Probability Distributions

As stated in the introduction, the square map can be considered as a vectorial Boolean function from the set of *n*-bit natural numbers to the set of *2n*-bit natural numbers; more precisely, we consider the function

$$f: F_2^n \to F_2^{2n}$$
$$x \mapsto y = f(x) = x^2.$$

In the rest of this section we study the probability distribution of the induced random variable y on F_2^{2n}. We introduce the indicator function for natural squares and from this the probability distribution of the output of the square map is computed.

For each $a \in F_2^{2n}$, we define

$$I(a) = \begin{cases} \lfloor \sqrt{a} \rfloor - \lfloor \sqrt{a-1} \rfloor & a \neq 0, \\ 1 & a = 0. \end{cases}$$

In fact, if a is a square then we have $I(a) = 1$ and otherwise, $I(a) = 0$. So,

$$P(y = a) = \frac{I(a)}{2^n}.$$

It is not hard to see that for $a < b$, we have

$$\sum_{i=a}^{b} P(y = i) = \begin{cases} \dfrac{\lfloor \sqrt{b} \rfloor - \lfloor \sqrt{a-1} \rfloor}{2^n} & a \neq 0, \\ \dfrac{\lfloor \sqrt{b} \rfloor + 1}{2^n} & a = 0. \end{cases}$$

Theorem 3.1. For each $1 \le t \le 2n$ and for every $a = (a_{t-1}, \dots, a_0) \in F_2^t - \{\mathbf{0}\}$,
we have

$$P(y_{t-1} = a_{t-1}, \dots, y_0 = a_0) = \frac{1}{2^n} \sum_{j=0}^{2^{2n-t}-1} \left(\left\lfloor \sqrt{j2^t + a} \right\rfloor - \left\lfloor \sqrt{j2^t + a - 1} \right\rfloor \right),$$

and

$$P(y_{t-1} = 0, \dots, y_0 = 0) = \tfrac{1}{2^n}\left(1 + \textstyle\sum_{j=1}^{2^{2n-t}-1} \left(\lfloor \sqrt{j2^t} \rfloor - \lfloor \sqrt{j2^t - 1} \rfloor \right)\right). \quad (1)$$

Proof: We have

$$P(y_{t-1} = a_{t-1}, \dots, y_0 = a_0) = \sum_{j=0}^{2^{2n-t}-1} P(y = j2^t + a) = \frac{1}{2^n} \sum_{j=0}^{2^{2n-t}-1} \left(\left\lfloor \sqrt{j2^t + a} \right\rfloor - \left\lfloor \sqrt{j2^t + a - 1} \right\rfloor \right),$$

and (1) is proved in the same manner.

Now, we study the probability distribution of the upper bits of the square map.

Theorem 3.2. For each $1 \le t \le 2n$ and for every $a = (a_{t-1}, \dots, a_0) \in F_2^t - \{\mathbf{0}\}$,
we have

$$P(y_{2n-1} = a_{t-1}, \dots, y_{2n-t} = a_0) = \frac{\left\lfloor \sqrt{(a+1)2^{2n-t} - 1} \right\rfloor - \left\lfloor \sqrt{a2^{2n-t} - 1} \right\rfloor}{2^n},$$

and,

$$P(y_{2n-1} = 0, \dots, y_{2n-t} = 0) = \tfrac{1}{2^n}\left(1 + \left\lfloor \sqrt{2^{2n-t} - 1} \right\rfloor\right). \quad (2)$$

Proof: We have

$$P(y_{2n-1} = a_{t-1}, \dots, y_{2n-t} = a_0)$$
$$= \sum_{i=a2^{2n-t}}^{(a+1)2^{2n-t}-1} P(y = i)$$

$$= \frac{\left\lfloor\sqrt{(a+1)2^{2n-t}-1}\right\rfloor - \left\lfloor\sqrt{a2^{2n-t}-1}\right\rfloor}{2^n},$$

and (2) is proved in the same manner.

The proof of the following result is left to the reader.

Result 3.3. For each t and for every

$$a = (a_{t-1}, \dots, a_0) \in F_2^t - \{\mathbf{0}\},$$

we have

$$\lim_{n\to\infty} P(y_{2n-1} = a_{t-1}, \dots, y_{2n-t} = a_0) = \frac{\sqrt{a+1}-\sqrt{a}}{\sqrt{2^t}},$$

and

$$\lim_{n\to\infty} P(y_{2n-1} = 0, \dots, y_{2n-t} = 0) = \frac{1}{\sqrt{2^t}}.$$

Now we find the probability distribution of the t-th bit of the output of the square map.

Theorem 3.4. For each $0 \le t < 2n$ we have

$$P(y_t = 0) = \frac{1 + \sum_{j=0}^{2^{2n-t-1}-1}\left\lfloor\sqrt{(2j+1)2^t-1}\right\rfloor - \sum_{j=1}^{2^{2n-t-1}-1}\left\lfloor\sqrt{j2^{t+1}-1}\right\rfloor}{2^n}.$$

Proof: We have

$$P(y_t = 0) = \sum_{j=0}^{2^{2n-t-1}-1}\sum_{i=0}^{2^t-1} P(y = j2^{t+1} + i)$$

$$= \frac{\left\lfloor\sqrt{2^t-1}\right\rfloor + 1 + \sum_{j=1}^{2^{2n-t-1}-1}\left(\left\lfloor\sqrt{(2j+1)2^t-1}\right\rfloor - \left\lfloor\sqrt{j2^{t+1}-1}\right\rfloor\right)}{2^n}$$

$$= \frac{\left\lfloor\sqrt{2^t-1}\right\rfloor + 1 - \left\lfloor\sqrt{2^t-1}\right\rfloor}{2^n} + \frac{\sum_{j=0}^{2^{2n-t-1}-1}\left\lfloor\sqrt{(2j+1)2^t-1}\right\rfloor - \sum_{j=1}^{2^{2n-t-1}-1}\left\lfloor\sqrt{j2^{t+1}-1}\right\rfloor}{2^n}$$

$$= \frac{1 + \sum_{j=0}^{2^{2n-t-1}-1}\left\lfloor\sqrt{(2j+1)2^t-1}\right\rfloor - \sum_{j=1}^{2^{2n-t-1}-1}\left\lfloor\sqrt{j2^{t+1}-1}\right\rfloor}{2^n}.$$

By doing some tedious computations, we have obtained the probability distribution of the upper bit, the two upper bits and the three upper bits of the output of square map. For example, considering the correspondence between F_2^n and Z_{2^n}, we have

$$P(y_{2n-1} = 0, y_{2n-2} = 0) = \frac{|\{x \in Z_{2^n} | x^2 < 2^{2n-2}\}|}{2^n} = \frac{1}{2}.$$

The results are as follows:

$$P(y_{2n-1} = 0) = \frac{\left\lfloor\sqrt{2^{2n-1}}\right\rfloor + 1}{2^n},$$
$$P(y_{2n-1} = 1) = \frac{2^n - \left\lfloor\sqrt{2^{2n-1}}\right\rfloor - 1}{2^n},$$

and

$$P(y_{2n-1} = 0, y_{2n-2} = 0) = \frac{1}{2},$$
$$P(y_{2n-1} = 0, y_{2n-2} = 1) = \frac{\left\lfloor\sqrt{2}.\,2^{n-1}\right\rfloor + 1 - 2^{n-1}}{2^n},$$
$$P(y_{2n-1} = 1, y_{2n-2} = 0) = \frac{\left\lfloor\sqrt{3}.\,2^{n-1}\right\rfloor - \left\lfloor\sqrt{2}.\,2^{n-1}\right\rfloor}{2^n},$$
$$P(y_{2n-1} = 1, y_{2n-2} = 1) = \frac{2^n - 1 - \left\lfloor\sqrt{3}.\,2^{n-1}\right\rfloor}{2^n},$$

and

$$P(y_{2n-1} = 0, y_{2n-2} = 0, y_{2n-3} = 0) = \frac{\left\lfloor\sqrt{2^{2n-3}}\right\rfloor + 1}{2^n},$$
$$P(y_{2n-1} = 0, y_{2n-2} = 0, y_{2n-3} = 1) = \frac{2^{n-1} - 1 - \left\lfloor\sqrt{2^{2n-3}}\right\rfloor}{2^n},$$
$$P(y_{2n-1} = 0, y_{2n-2} = 1, y_{2n-3} = 0) = \frac{\left\lfloor\sqrt{3.2^{2n-3}-1}\right\rfloor - 2^{n-1} + 1}{2^n},$$
$$P(y_{2n-1} = 0, y_{2n-2} = 1, y_{2n-3} = 1) = \frac{\left\lfloor\sqrt{2}.\,2^{n-1}\right\rfloor - \left\lfloor\sqrt{3.2^{2n-3}}\right\rfloor}{2^n},$$
$$P(y_{2n-1} = 1, y_{2n-2} = 0, y_{2n-3} = 0) = \frac{\left\lfloor\sqrt{5.2^{2n-3}-1}\right\rfloor - \left\lfloor\sqrt{2}.\,2^{n-1}\right\rfloor}{2^n},$$
$$P(y_{2n-1} = 1, y_{2n-2} = 0, y_{2n-3} = 1) = \frac{\left\lfloor\sqrt{3.2^{2n-2}-1}\right\rfloor - \left\lfloor\sqrt{5.2^{2n-3}}\right\rfloor}{2^n},$$
$$P(y_{2n-1} = 1, y_{2n-2} = 1, y_{2n-3} = 0) = \frac{\left\lfloor\sqrt{7.2^{2n-3}-1}\right\rfloor - \left\lfloor\sqrt{3}.\,2^{n-1}\right\rfloor}{2^n},$$
$$P(y_{2n-1} = 1, y_{2n-2} = 1, y_{2n-3} = 1) = \frac{2^n - 1 - \left\lfloor\sqrt{7.2^{2n-3}}\right\rfloor}{2^n}.$$

It is worth noting that, at first glance it seems that these distributions are not equal to the results of Theorem 3-2, but actually they are: we have verified the equality of these formulas. For instance, the previous computations state that

$$P(y_{2n-1} = 1, y_{2n-2} = 1, y_{2n-3} = 1) = \frac{2^n - 1 - \left\lfloor\sqrt{7.2^{2n-3}-1}\right\rfloor}{2^n},$$

and based on Theorem 3-2, we have

$$P(y_{2n-1}=1, y_{2n-2}=1, y_{2n-3}=1)$$

$$=\frac{\left\lfloor\sqrt{(7+1)2^{2n-3}-1}\right\rfloor-\left\lfloor\sqrt{7.2^{2n-3}-1}\right\rfloor}{2^n}$$
$$=\frac{\left\lfloor\sqrt{2^{2n}-1}\right\rfloor-\left\lfloor\sqrt{7.2^{2n-3}-1}\right\rfloor}{2^n}.$$

Now, it is not hard to verify that $\left\lfloor\sqrt{2^{2n}-1}\right\rfloor = 2^n-1$ and $\left\lfloor\sqrt{7.2^{2n-3}-1}\right\rfloor=\left\lfloor\sqrt{7.2^{2n-3}}\right\rfloor$.

4. The Imbalance of the Square and the Square Root Maps

In this section, we introduce a measure for computing the imbalance of maps, and based upon this criterion, the imbalance of the square map is computed. Then, a new map called the square root map is introduced and the imbalance of this new map is, computed too.

Definition 4.1. (Cover and Thomas, Chap. 11, 2006): Suppose that P_1 and P_2 are two probability distributions on a finite sample space $\mathcal{X}$. The distance between these two probability distributions is defined as

$$D(P_1,P_2)=\sum_{x\in\mathcal{X}}|P_1(x)-P_2(x)|.$$

Now, let n, m and d be natural numbers. For a function $f:A\to B$ with $|A|=n$, $|B|=m$ and $n=dm$, we define the probability distribution P_1 on B as

$$P_1(b)=\frac{|f^{-1}(b)|}{n},\qquad b\in B,$$

and we define the probability distribution P_2 on B as the uniform distribution:

$$P_2(b)=\frac{d}{n},\qquad b\in B.$$

Definition 4.2. (Dehnavi, et al. 2013): We define a criterion for measuring the imbalance D_f for the function $f:A\to B$, with $|A|=n$, $|B|=m$ and $n=dm$, as

$$D_f=\frac{m}{2(m-1)}D(P_1,P_2)=\frac{\sum_{b\in B}\left||f^{-1}(b)|-d\right|}{2(m-1)d}.$$

Lemma 4.3. (Dehnavi, et al. 2013): For each function $f:A\to B$ with $|A|=n$, $|B|=m$ and $n=dm$, we have

$$0\le D_f\le 1;$$

further, for each balanced function we have $D_f=0$ and for every constant function we have $D_f=1$. The proof of the next lemma is not hard.

Lemma 4.4. Let t be a fixed natural number. For the real function

$$f:\{0,1,\dots,2^n-1\}\to\mathbb{R},$$
$$x\mapsto f(x)=\frac{\sqrt{x+1}-\sqrt{x}}{\sqrt{2^t}}-\frac{1}{2^t},$$

we have

$$\begin{cases} f(x)>0 & x<2^{t-2},\\ f(x)<0 & x\ge 2^{t-2}.\end{cases}$$

Based on Lemma 4-4 and Result 3-3, we can obtain the imbalance of the limit probability distribution of the upper t bits of the output of square map.

Theorem 4.5. Let $f:F_2^n\to F_2^{2n}$ be defined as $f(x)=x^2$; let f_t be the limit distribution of vectorial Boolean function $(f_{2n-1},f_{2n-2},\dots,f_{2n-t})$ for a fixed t. Then,

$$D_{f_t}=\frac{2^{t-2}}{2^t-1}.$$

Proof: We have

$$D_{f_t}=\frac{2^t}{2(2^t-1)}\left(\sum_{a=0}^{2^{t-2}-1}\left(\frac{\sqrt{a+1}-\sqrt{a}}{\sqrt{2^t}}-\frac{1}{2^t}\right)+\sum_{a=2^{t-2}}^{2^t-1}\left(\frac{1}{2^t}-\frac{\sqrt{a+1}-\sqrt{a}}{\sqrt{2^t}}\right)\right)$$
$$=\frac{2^t}{2(2^t-1)}\left(\left(\frac{\sqrt{2^{t-2}}}{\sqrt{2^t}}-\frac{2^{t-2}}{2^t}\right)+\left(\frac{3\times 2^{t-2}}{2^t}-\frac{\sqrt{2^t}-\sqrt{2^{t-2}}}{2^t}\right)\right)$$
$$=\frac{2^t}{2(2^t-1)}\left(\frac{2\times 2^{t-2}}{2^t}+\frac{2\sqrt{2^{t-2}}-\sqrt{2^t}}{\sqrt{2^t}}\right)$$
$$=\frac{2^{t-2}}{2^t-1}.$$

Now, we introduce the square root map and compute the probability distribution of the component Boolean functions of the output of this map along with the imbalance of it. The proof of the next lemma is easy.

Lemma 4-6: Suppose that $f:F_2^{2n}\to F_2^n$ is defined as $f(x)=\lfloor\sqrt{x}\rfloor$. Then for each $0\le a<2^n$, we have

$$|f^{-1}(a)| = 2a + 1.$$

Now, we can obtain the imbalance of the square root map.

Theorem 4.7. Suppose that $f: F_2^{2n} \to F_2^n$ is defined as $f(x) = \lfloor \sqrt{x} \rfloor$. Then,

$$D_f = \frac{2^{n-2}}{2^n - 1}.$$

Proof: We have

$$\begin{aligned} D_f &= \frac{\sum_{i=0}^{2^n-1} \left| |f^{-1}(i)| - 2^n \right|}{2^{n+1}(2^n - 1)} \\ &= \frac{1}{2^{n+1}(2^n - 1)} \left(\sum_{i=0}^{2^{n-1}-1} (2^n - (2i+1)) \right) \\ &\quad + \sum_{i=2^{n-1}}^{2^n-1} (2i + 1 - 2^n) \\ &= \frac{2^{n-2}}{2^n - 1}. \end{aligned}$$

The probability distribution of the component Boolean functions of the square root map shall be obtained in the next theorem.

Theorem 4.8. Suppose that $f: F_2^{2n} \to F_2^n$ is defined as $y = f(x) = \lfloor \sqrt{x} \rfloor$. Then,

$$P(y_t = 0) = \frac{1}{2} - \frac{1}{2^{n-t+1}}.$$

Proof: We have

$$\begin{aligned} P(y_t = 0) &= \sum_{j=0}^{2^{n-t-1}-1} \sum_{i=0}^{2^t-1} P(y = j2^{t+1} + i) \\ &= \frac{1}{2^{2n}} \sum_{j=0}^{2^{n-t-1}-1} \sum_{i=0}^{2^t-1} (j2^{t+2} + 2i + 1) \\ &= \frac{1}{2} - \frac{1}{2^{n-t+1}}. \end{aligned}$$

Note 4.9. Suppose that $f: F_2^{2n} \to F_2^n$ is defined as $y = f(x) = \lfloor \sqrt{x} \rfloor$ and $g: F_2^{2n} \to F_2^n$ is defined as $z = g(x, y) = xy \ mod \ 2^n$.
Then, based on (Dehnavi et al. 2013), we have

$$P(y_t = 0) = P(z_{n-1-t} = 1),$$

and

$$D_f = D_g.$$

In spite of the fact that probability distributions of the operator of multiplication modulo, a power of two and the square root map are different, the probability distribution of their component Boolean functions are closely related and their imbalance are equal.

References

Boesgaard, M., Vesterager, M., Pedersen, T., Christiansen, J., & Scavenius, O. (2003). Rabbit: A New High-Performance Stream Cipher, in *Fast Software Encryption (FSE'03)*, LNCS 2887, 307–329, Springer-Verlag.

Cover, T. M., & Thomas, J. A. (2003). *Elements of Information Theory*. Second Edition, John Wiley & Sons.

Dehnavi, S. M., Mahmoodi Rishakani, A., Mirzaee Shamsabad, M. R., & Pasha, E. (2013). Cryptographic Properties of the Operator of Multiplication Modulo a Power of Two. *Journal of Science, Kharazmi University*. 12(1), 327–338 (In Persian).

Rivest, R. L., Robshaw, M. J. B., Sidney, R., & Yin, Y. L. (1998). The RC6 Block Cipher. *Proceeding of 1st Advanced Encryption Standard Candidate Conference*, Venture, California, Aug. 20–22.

Stinson, D. R. (2003). *Cryptography-Theory and Practice*, 3rd edn. Chapman & Hall/CRC, Boca Raton.