NPCR and UACI Randomness Tests for Image Encryption

Yue Wu, *Student Member, IEEE*, Joseph P. Noonan, *Life Member, IEEE*, and Sos Agaian, *Senior Member, IEEE*

***Abstract*—The number of changing pixel rate (NPCR) and the unified averaged changed intensity (UACI) are two most common quantities used to evaluate the strength of image encryption algorithms/ciphers with respect to differential attacks. Conventionally, a high NPCR/UACI score is usually interpreted as a high resistance to differential attacks. However, it is not clear how high NPCR/UACI is such that the image cipher indeed has a high security level. In this paper, we approach this problem by establishing a mathematical model for ideally encrypted images and then derive expectations and variances of NPCR and UACI under this model. Further, these theoretical values are used to form statistical hypothesis NPCR and UACI tests. Critical values of tests are consequently derived and calculated both symbolically and numerically. As a result, the question of whether a given NPCR/UACI score is sufficiently high such that it is not discernible from ideally encrypted images is answered by comparing actual NPCR/UACI scores with corresponding critical values. Experimental results using the NPCR and UACI randomness tests show that many existing image encryption methods are actually not as good as they are purported, although some methods do pass these randomness tests.**

***Index Terms*—Differential Attacks, Randomness Test, Image Encryption, UACI, NPCR**

I. Introduction

Differntial attack/cryptanalysis is a general name of attacks/cryptanalysis applicable primarily to block ciphers working on binary sequences. The discovery of differential cryptanalysis is usually attributed to Eli Biham and Adi Shamir, who published papers [1, 2] about this type of attacks to various ciphers, including a theoretical weakness of the Data Encryption Standard (DES) [3]. Since then, the differential attack becomes a common attack that has to be considered during the cipher design.

Manuscript received March 29, 2011. Manuscript accepted April 26, 2011. This research was supported by the Department of Electrical and Computer Engineering, Tufts University, MA.

Yue Wu is with the Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155 USA (phone: 617-627-3217; fax: 617-627-3220; e-mail: ywu03@ece.tufts.edu).

Joseph P. Noonan is with the Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155 USA. (e-mail: jnoonan@ece.tufts.edu).

Sos Agaian is with the Department of Electrical and Computer Engineering, Tufts University, MA 02155 USA. He is also with the Department of Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, TX 78249 USA. (email: Sos.Agaian@utsa.edu)

In binary sequence encryption, the cipher resistance to differential attacks is normally analyzed directly via calculating the independence matrix [4] between any two output bits and the dependence matrix [4] between the input bits and output bits. However, unlike binary sequence encryption, image encryption [5-14] is a relatively new area with distinctive characteristics including 1) it is a type of two-dimensional data with high information redundancy [15]; and 2) it usually contains of a large number of pixels, each of which is composed of a number of binary bits. All these properties make the conventional ciphers designed for binary data inappropriate for image data [15]. For the same reason, randomness tests for binary data are also not appropriate for image encryption methods/ciphers.

In image encryption, the cipher resistance to differential attacks is commonly analyzed via the NPCR and UACI tests [5-14]. The NPCR and UACI are designed to test the number of changing pixels and the number of averaged changed intensity between ciphertext images, respectively, when the difference between plaintext images is subtle (usually a single pixel). Although these two tests are compactly defined and are easy to calculate, test scores are difficult to interpret in the sense of whether the performance is good enough. For example, the upper-bound of the NPCR score is 100%, and thus it is believed that the NPCR score of a secure cipher should be very close to this upper-bound. However, the question is how close is 'close'? A NPCR score of 99% is close or a score of 99.9% or neither of them is close enough. Therefore, it is trivial to answer the quantitative question that what are the NPCR and UACI scores for one image encryption algorithm/cipher, without knowing the answer of the qualitative question that whether this algorithm/cipher is able to generate secure enough ciphertext with resistance to differential attacks.

Inspired by the FIPS 140-1 [16] and its successor FIPS 140-2 [17] randomness test sets for binary ciphers, we believed that randomness tests giving qualitative results rather than pure quantitative results should be derived for image encryption as well. In this paper, we focus on the NPCR and UACI tests and give our solutions to answer the qualitative question about NPCR and UACI tests for image encryption.

The remainder of the paper is organized as follows: Section II gives the mathematical model of an ideally encrypted image and derives the expectations, variances and hypothesis tests of NPCR and UACI; Section III gives numerical results of these expectations, variances and lookup tables of critical values for

hypothesis tests; Section IV shows results of the proposed randomness tests of NPCR and UACI for a number of published image encryption methods; Section V concludes the paper and discusses our future work

II. Mathematical Derivations of NPCR and UACI Randomness Tests

A. NPCR and UACI Definitions

For our best knowledge, NPCR and UACI are first shown in 2004 [5, 18], both of which point to Yaobin Mao and Guanrong Chen. Since then NPCR and UACI become two widely used security analyses in the image encryption community for differential attacks.

Suppose ciphertext images before and after one pixel change in a plaintext image are C^1 and C^2, respectively; the pixel value at grid (i,j) in C^1 and C^2 are denoted as $C^1(i,j)$ and $C^2(i,j)$; and a bipolar array D is defined in Eqn. (1). Then the NPCR and UACI can be mathematically defined by Eqns. (2) and (3), respectively, where symbol T denotes the total number pixels in the ciphertext, symbol F denotes the largest supported pixel value compatible with the ciphertext image format, and $|.|$ denotes the absolute value function.

$$D(i,j) = \begin{cases} 0, if\, C^1(i,j) = C^2(i,j) \\ 1, if\, C^1(i,j) \neq C^2(i,j) \end{cases} \quad (1)$$

$$\text{NPCR: } \mathcal{N}(C^1, C^2) = \sum_{i,j} \frac{D(i,j)}{T} \times 100\% \quad (2)$$

$$\text{UACI: } \mathcal{U}(C^1, C^2) = \sum_{i,j} \frac{|C^1(i,j) - C^2(i,j)|}{F \cdot T} \times 100\% \quad (3)$$

It is clear that NPCR concentrates on the absolute number of pixels which changes value in differential attacks, while the UACI focuses on the averaged difference between two paired ciphertext images.

The range of NPCR is $[0,1]$. When $\mathcal{N}(C^1, C^2) = 0$, it implies that all pixels in C^2 remain the same values as in C^1. When $\mathcal{N}(C^1, C^2) = 1$, it implies that all pixel values in C^2 are changed compared to those in C^1. In other words, it is very difficult to establish relationships between this pair of ciphertext image C^1 and C^2. However, $\mathcal{N}(C^1, C^2) = 1$ rarely happens, because even two independently generated true random images fail to achieve this NPCR maximum with a high possibility, especially when the image size is fairly large compared to F.

The range of UACI is clearly $[0,1]$ as well, but it is not obvious that what a desired UACI for two ideally encrypted images is. Fortunately, these results will be given in next sections with the form of expectations and variances.

B. Ideally Encrypted Image

Before start to derive the interested statistics about NPCR and UACI for ideally encrypted images, the term of 'ideally encrypted image' has to be clarified first. Although it may be considered differently in other literature, in this paper, we consider an ideally encrypted image is some image that cannot be discernible from a true random image. More specifically,

Definition 1. *Ideally Encrypted Image*

An ideally encrypted image I is a random field at size of M-by-N, where for any fixed integer $i \in [1,M]$ and $j \in [1,N]$, the random variable of pixel value $I(i,j)$ identically and independently (i.i.d) follows a discrete uniform distribution on 0 to I's largest supported integer F, i.e. $\forall i \in [1,M]$, $j \in [1,N]$, $\exists\, I(i,j) \sim i.i.d.\, \mathbb{U}(0,F)$.

It is noticeable that the above definition is plausible in the context of image encryption, where the aim of encryption is to obtain random-like ciphertext images such that attackers cannot figure out the internal relations between plaintext and ciphertext. In fact, other security analyses [5-14], e.g. histogram analysis, entropy analysis and autocorrelation analysis, are all designed to test whether or not a ciphertext image is random-like.

For any pixel at any location in an ideally encrypted image I, its value is equally likely to be an arbitrary intensity level l in $[0,F]$, namely $\Pr[I(i,j) = l] = 1/(F+1)$. In order to save notations, the spatial index (i,j) can be expressed by an absolutely index k as Eqn. (4) shows. As a result, we have $D(i,j) = D(k)$.

$$k = (j-1)N + i \quad (4)$$

C. NPCR Test

In this section, the expectation and the variance of NPCR for two ideally encrypted images are calculated first and then an α-level hypothesis test is derived based on these two statistics. For simplicity,

Theorem I. *For the xth pixels ($x \in [1,MN]$) in two ideally encrypted images defined in Definition 1, define a random variable*

$$d = \begin{cases} 0, if\, I^1(x) = I^2(x) \\ 1, if\, I^1(x) \neq I^2(x) \end{cases}$$

Then this random variable d follows a Bernoulli distribution with the parameter $p = F/(F+1)$.

Proof. Using the assumption of independence and $\mathbb{U}(0,F)$, it is easy to see,

$$\begin{aligned} \Pr[d=0] &= \Pr[I^1(x) = I^2(x)] \\ &= \sum_{l=0}^{F} \Pr[I^1(x) = l \mid I^2(x) = l] \cdot \Pr[I^2(x) = l] \\ &= \sum_{l=0}^{F} \Pr[I^1(x) = l] \cdot \Pr[I^2(x) = l] \\ &= 1/(F+1) \end{aligned}$$

Consequently, $\Pr[d=1] = 1 - \Pr[d=1] = F/(F+1)$. Therefore, $d \sim \mathbb{B}(F/(F+1))$. ■

Moreover, if the total number of pixels whose $D(x) = 1$ is denoted as a random variable S, then S has the Binomial distribution as Theorem II states.

Theorem II. *The random variable $S = \sum_{x=1}^{MN} D(x)$ defined on two ideally encrypted images follows a Binomial distribution $\mathbb{B}(MN, p)$, where $p = F/(F+1)$.*

Proof. Using the conclusion of Theorem I and i.i.d property between pixels, it is clear that

$$\Pr[S = k] = \binom{MN}{k}\left(\frac{F}{F+1}\right)^{k}\left(\frac{1}{F+1}\right)^{MN-k}$$

which is the Binomial distribution $\mathbb{B}(MN, p)$. ■

Therefore, the expectation and the variance of S are explicitly defined as Eqns. (5) and (6), respectively.

$$\mu_S = MN \cdot p = MNF/(F+1) \tag{5}$$

$$\sigma_S^2 = MN \cdot p \cdot (1-p) = MNF/(F+1)^2 \tag{6}$$

It is clear that this random variable S is a scaled version of the NPCR score, where $\mathcal{N}(I^1, I^2) = \sum_{i,j} D(i,j)/T = S/MN$. Therefore, $\mathcal{N}(I^1, I^2) \sim \mathbb{B}(MN, p)$, if two test ciphertext images I^1 and I^2 of size M-by-N are ideally encrypted. That is

$$\Pr\left[\mathcal{N}(I^1, I^2) = \frac{k}{MN}\right] = \binom{MN}{k}\left(\frac{F}{F+1}\right)^{k}\left(\frac{1}{F+1}\right)^{MN-k} \tag{7}$$

$$\mu_{\mathcal{N}} = \mu_S/MN = F/(F+1) \tag{8}$$

$$\sigma_{\mathcal{N}}^2 = \frac{\sigma_S^2}{(MN)^2} = \frac{F}{MN(F+1)^2} \tag{9}$$

As a result, the following statistical test can be used as a test of NPCR for image encryption:

Definition 2. *Randomness Test for NPCR*
Suppose C^1 and C^2 are two test ciphertext images at the size M-by-N, the hypotheses with α-level significance for $\mathcal{N}(C^1, C^2)$, then, are

$$\begin{cases}\mathcal{H}_0: \mathcal{N}(C^1, C^2) = \mu_{\mathcal{N}} \\ \mathcal{H}_1: \mathcal{N}(C^1, C^2) < \mu_{\mathcal{N}}\end{cases}$$

where we reject $\mathcal{H}_0$, when $\mathcal{N}(C^1, C^2) < \mathcal{N}^*$, the critical value of the NPCR test; otherwise we accept $\mathcal{H}_0$. The critical value $\mathcal{N}^*$ is defined in Eqn. (10), where $\Phi^{-1}(.)$ is the inverse cumulative density function (CDF) of the standard Normal distribution $\mathbb{N}(0,1)$.

$$\begin{aligned}\mathcal{N}_\alpha^* &= \mu_{\mathcal{N}} - \Phi^{-1}(\alpha)\sigma_{\mathcal{N}} \\ &= \left(F - \Phi^{-1}(\alpha)\sqrt{\frac{F}{MN}}\right)\Big/(F+1)\end{aligned} \tag{10}$$

D. UACI Test for Ideally Encrypted Image

Similarly to NPCR test, the UACI test derived in this section is also with respect to two ideally encrypted images.

Consider a new random field A, which is the absolute difference between C^1 and C^2 as Eqn. (11) shows. Since pixel values in C^1 and C^2 are both i.i.d, pixels in A is also i.i.d with some unknown distribution.

$$A = |C^1 - C^2| \tag{11}$$

Let $a = A(x) = |I^1(x) - I^2(x)|$, then this random variable for the averaged changed intensity for one pixel location in two ideally encrypted images follows a discrete distribution showed in Theorem III.

Theorem III. *If $a = A(x) = |I^1(x) - I^2(x)|$, which is the changed intensity of two ideally encrypted images at location x, then*

$$\Pr[a = k] = \begin{cases} 1/(F+1) & if\ k = 0 \\ 2(F+1-k)/(F+1)^2 & 'if\ k \in (0, F] \end{cases}$$

Proof. From Theorem I, it is clear that when $k = 0$

$$\Pr[a = 0] = \Pr[d = 0] = 1/(F+1)$$

When $k \in (0, F+1]$,

$$\begin{aligned}\Pr[a = k] &= \Pr[|I^1(x) - I^2(x)|] \\ &= \Pr[I^1(x) - I^2(x) = k] + \Pr[I^2(x) - I^1(x) = k]\end{aligned}$$

Calculate $\Pr[I^1(x) - I^2(x) = k]$ using Definition 1, we obtain

$$\begin{aligned}\Pr[I^1(x) - I^2(x) = k] &= \sum_{l=0}^{F} \Pr[I^1(x) = l] \cdot \Pr[I^2(x) = l - k] \\ &= \sum_{l=k}^{F} \Pr[I^1(x) = l] \cdot \Pr[I^2(x) = l - k] \\ &= (F - k + 1)/(F+1)^2\end{aligned}$$

Similarly, $\Pr[I^2(x) - I^1(x) = k] = (F - k + 1)/(F+1)^2$.
Thus, $\Pr[a = k] = 2(F + 1 - k)/(F+1)^2$. ■

Theorem III gives the probability density function (PDF) of the random variable a and the i.i.d distribution in the random field A as well. In addition, the mean and the variance of a can also be obtained as Eqns. (12) and (13) show.

$$\mu_a = \sum_{k=1}^{F} 2k(F+1-k)/(F+1)^2 = \frac{F(F+2)}{3F+3} \tag{12}$$

$$\begin{aligned}\sigma_a^2 &= \sum_{k=1}^{F} 2k^2(F+1-k)/(F+1)^2 - \mu_a^2 \\ &= \frac{F(F+2)}{6} - \left(\frac{F(F+2)}{3F+3}\right)^2 \\ &= \frac{F(F+2)(F^2+2F+3)}{18(F+1)^2}\end{aligned} \tag{13}$$

Let quantity $B = \sum_{x=1}^{MN} |I^1(x) - I^2(x)|/MN$, then this B is nothing but the mean value of A, as Eqn. (14) shows. Moreover the relationship between B and UACI is $\mathcal{U}(C^1, C^2) = B/F$, which implies B is a scaled version of the UACI score.

$$B = \sum_{x=1}^{MN} A(x)/MN \tag{14}$$

Theorem IV. If $B = \sum_{x=1}^{MN} |I^1(x) - I^2(x)| / MN$ is the scaled version of UACI between two ideally encrypted images I^1 and I^2 whose plaintext images are slightly different, then $B \sim \mathbb{N}(\mu_a, \sigma_a^2/MN)$

Proof.

The Central Limit Theorem (CLT) tells that as long as the sample size n is large enough, the sample mean of any i.i.d distributed sample with an arbitrary PDF with an average μ and a finite σ^2 is approximately a Gaussian $\mathbb{N}(\mu, \sigma^2/n)$. In our case, n is the number of pixels and is usually much large than 100, which is the sample size believed the CLT can be applied [19, 20].

Because $\forall x$, $A(x)$ are i.i.d distributed with PDF specified in Theorem III. Therefore, $B \sim \mathbb{N}(\mu_B, \sigma_B^2)$ where μ_B and σ_B^2 are shown in Eqns.(15) and (16), respectively. ■

$$\mu_B = \mu_a = \frac{F(F+2)}{3F+3} \tag{15}$$

$$\sigma_B^2 = \frac{\sigma_a^2}{MN} = \frac{F(F+2)(F^2+2F+3)}{18(F+1)^2 MN} \tag{16}$$

As a result, we obtain the expectation and the variance for the UACI test as follows:

$$\mu_{\mathcal{U}} = \mu_B / F = (F+2)/(3F+3) \tag{17}$$

$$\sigma_{\mathcal{U}}^2 = \sigma_B^2 / F^2 = \frac{(F+2)(F^2+2F+3)}{18(F+1)^2 MNF} \tag{18}$$

Since the reference results have been derived from the ideally encrypted image, the following statistical test can be used to test UACI:

Definition 3. *Randomness Test for UACI*

Suppose C^1 and C^2 are two test ciphertext images at the size M-by-N, then the hypotheses with α-level significance for $\mathcal{U}(C^1, C^2)$, then, are

$$\begin{cases} \mathcal{H}_0 : \mathcal{U}(C^1, C^2) = \mu_{\mathcal{U}} \\ \mathcal{H}_1 : \mathcal{U}(C^1, C^2) \neq \mu_{\mathcal{U}} \end{cases}$$

where we reject $\mathcal{H}_0$, when $\mathcal{U}(C^1, C^2) \notin (\mathcal{U}_\alpha^{*-}, \mathcal{U}_\alpha^{*+})$, the critical values of the NPCR test; otherwise we accept $\mathcal{H}_0$. The critical value $\mathcal{U}_\alpha^{*-}$ and $\mathcal{U}_\alpha^{*+}$ are defined in Eqns. (19) and (20), respectively, where $\Phi^{-1}(.)$ is the inverse CDF of the standard Normal distribution $\mathbb{N}(0,1)$.

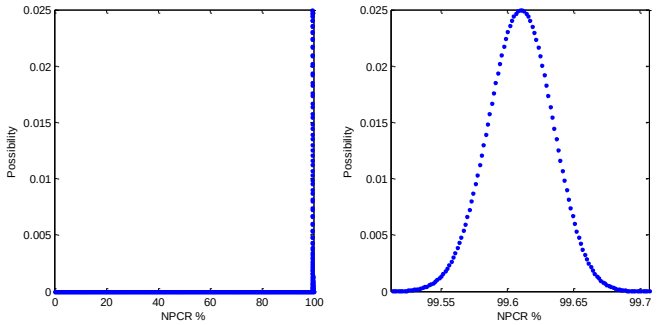
$$\mathcal{U}_\alpha^{*-} = \mu_{\mathcal{U}} - \Phi^{-1}(\alpha/2)\sigma_{\mathcal{U}} \tag{19}$$

$$\mathcal{U}_\alpha^{*+} = \mu_{\mathcal{U}} + \Phi^{-1}(\alpha/2)\sigma_{\mathcal{U}} \tag{20}$$

III. Numerical Results for NPCR and UACI Randomness Tests

A. Numerical Results for NPCR

As the previous section derived, the expectation, the variance and the PDF of NPCR statistic have already been shown in Eqns. (5)-(7). The distribution of the NPCR random variable $\mathcal{N}(I^1, I^2)$ for two true random images follows a Binomial distribution $\mathbb{B}(MN, p)$. When $MN = 256 \times 256$ and $F = 255$, this distribution is shown Fig. 1, where figure (b) is an enlarged version for the peak in figure (a). From Fig. 1, it is clear that $\mathcal{N}(I^1, I^2)$ has Gaussian-like distribution. Indeed, a Binomial distribution can be approximated as a Gaussian distribution whenever the condition $0 < \mu_S - 3\sigma_s < \mu_S + 3\sigma_s < MN$ is satisfied [21].



(a) PDF NPCR $f_{\mathcal{N}}$ (b) Zoom-in $f_{\mathcal{N}}$

Fig. 1. PDF of NPCR for $M = N = 256$ and $F = 255$

Numerical results of NPCR critical values with respect to different parameter combinations are given in Table I. From Eqns. (5) and (6), it is noticeable that $\mu_{\mathcal{N}}$ is a constant and $\sigma_{\mathcal{N}}$ is proportional to $1/\sqrt{MN}$, respectively, when F is fixed. Therefore, as the $M \times N$ increases four times, $\mu_{\mathcal{N}}$ remains unchanged, while $\sigma_{\mathcal{N}}$ deceases a half.

In Table I, $\mathcal{N}_{0.05}^*$, $\mathcal{N}_{0.01}^*$, and $\mathcal{N}_{0.001}^*$denote the critical values to reject the null hypothesis with respect to the significance level $\alpha = 0.05$, $\alpha = 0.01$ and $\alpha = 0.001$. This means that if $\mathcal{N}(C^1, C^2)$, the NPCR test for two paired ciphertext images C^1 and C^2, less than $\mathcal{N}_\alpha^*$, then C^1 and C^2 are *NOT* randomly-like with an α-level of significance. In other words, the possibility to say 'C^1 and C^2 are not random-like', when they are random-like, is α, which is a small quantity.

B. Numerical Results for UACI

Table II shows related numerical results for UACI. In this table, it is noticeable that $\mu_{\mathcal{U}}$ is independent of $M \times N$. Because $\mu_{\mathcal{U}} = (F+2)/(3F+3)$, which is a single variable function about F (see Eqn. (17)), the largest allowed integer related to the image format. Meanwhile, $\sigma_{\mathcal{U}}$ halves its value as MN increases in the table. This is because $\sigma_{\mathcal{U}}$ is proportional to $1/\sqrt{MN}$, whenever MN increases four times, $\sigma_{\mathcal{U}}$ halves itself.

Unlike the critical value $\mathcal{N}_\alpha^*$ for NPCR test, the critical value $\mathcal{U}_\alpha^*$for UACI test is composed of two parts, the left value $\mathcal{U}_\alpha^{*-}$ and the right value $\mathcal{U}_\alpha^{*+}$. All these values are listed in Table II.

For any tested $\mathcal{U}(C^1, C^2)$, if it is out of the acceptance interval $(\mathcal{U}_\alpha^{*-}, \mathcal{U}_\alpha^{*+})$, we reject the null hypothesis and say the tested ciphertext images C^1 and C^2 are *NOT* random-like. Again, this assertion maybe wrong, but the possibility to make a mistake is only α, which is a small quantity.

TABLE I. NUMERICAL RESULTS FOR NPCR RANDOMNESS TEST

	Binary Image: $F = 1$					Gray Image: $F = 255$				
$M \times N$	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$	$\mathcal{N}^*_{0.05}$	$\mathcal{N}^*_{0.01}$	$\mathcal{N}^*_{0.001}$	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$	$\mathcal{N}^*_{0.05}$	$\mathcal{N}^*_{0.01}$	$\mathcal{N}^*_{0.001}$
64×64	50.0000%	0.7813%	48.7150%	48.1825%	47.5858%	99.6094%	0.0975%	99.4491%	99.3826%	99.3082%
128×128	50.0000%	0.3906%	49.3575%	49.0913%	48.7929%	99.6094%	0.0487%	99.5292%	99.4960%	99.4588%
256×256	50.0000%	0.1953%	49.6787%	49.5456%	49.3964%	99.6094%	0.0244%	99.5693%	99.5527%	99.5341%
512×512	50.0000%	0.0977%	49.8394%	49.7728%	49.6982%	99.6094%	0.0122%	99.5893%	99.5810%	99.5717%
1024×1024	50.0000%	0.0488%	49.9197%	49.8864%	49.8491%	99.6094%	0.0061%	99.5994%	99.5952%	99.5906%

TABLE II. NUMERICAL RESULTS FOR UACI RANDOMNESS TEST

	Binary Image: $F = 1$					Gray Image: $F = 255$				
$M \times N$	μ_u	σ_u	$\mathcal{U}^{*-}_{0.05}/\mathcal{U}^{*+}_{0.05}$	$\mathcal{U}^{*-}_{0.01}/\mathcal{U}^{*+}_{0.01}$	$\mathcal{U}^{*-}_{0.001}/\mathcal{U}^{*+}_{0.001}$	μ_u	σ_u	$\mathcal{U}^{*-}_{0.05}/\mathcal{U}^{*+}_{0.05}$	$\mathcal{U}^{*-}_{0.01}/\mathcal{U}^{*+}_{0.01}$	$\mathcal{U}^{*-}_{0.001}/\mathcal{U}^{*+}_{0.001}$
64×64	50.0000%	0.7813%	48.4688%	47.9876%	47.4293%	33.4635%	0.3697%	32.7389%	32.5112%	32.2469%
			51.5312%	52.0124%	52.5707%			34.1882%	34.4159%	34.6802%
128×128	50.0000%	0.3906%	49.2344%	48.9938%	48.7146%	33.4635%	0.1849%	33.1012%	32.9874%	32.8552%
			50.7656%	51.0062%	51.2854%			33.8259%	33.9397%	34.0718%
256×256	50.0000%	0.1953%	49.6172%	49.4969%	49.3573%	33.4635%	0.0924%	33.2824%	33.2255%	33.1594%
			50.3828%	50.5031%	50.6427%			33.6447%	33.7016%	33.7677%
512×512	50.0000%	0.0977%	49.8086%	49.7485%	49.6787%	33.4635%	0.0462%	33.3730%	33.3445%	33.3115%
			50.1914%	50.2515%	50.3213%			33.5541%	33.5826%	33.6156%
1024×1024	50.0000%	0.0488%	49.9043%	49.8742%	49.8393%	33.4635%	0.0231%	33.4183%	33.4040%	33.3875%
			50.0957%	50.1258%	50.1607%			33.5088%	33.5231%	33.5396%

TABLE III. COMPARISON OF THEORETICAL VALUES AND EXPERIMENTAL VALUES

Binary Image: $F = 1$

	NPCR %				UACI %			
$M \times N$	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$	$\widehat{\mu_{\mathcal{N}}}$	$\widehat{\sigma_{\mathcal{N}}}$	μ_u	σ_u	$\widehat{\mu_u}$	$\widehat{\sigma_u}$
64×64	50.0000000000	0.7813000000	49.9984221458	0.7838076127	50.0000000000	0.7812500000	49.9984221458	0.7838076127
128×128	50.0000000000	0.3906000000	49.9944293455	0.3913540553	50.0000000000	0.3906250000	49.9944293455	0.3913540553
256×256	50.0000000000	0.1953000000	49.9965943224	0.1956158262	50.0000000000	0.1953125000	49.9965943224	0.1956158262
512×512	50.0000000000	0.0977000000	49.9988945723	0.0970774641	50.0000000000	0.0976562500	49.9988945723	0.0970774641
1024×1024	50.0000000000	0.0488000000	50.0011780387	0.0486855663	50.0000000000	0.0488281250	50.0011780387	0.0486855663

Gray Image: $F = 255$

	NPCR %				UACI %			
$M \times N$	$\mu_{\mathcal{N}}$	$\sigma_{\mathcal{N}}$	$\widehat{\mu_{\mathcal{N}}}$	$\widehat{\sigma_{\mathcal{N}}}$	μ_u	σ_u	$\widehat{\mu_u}$	$\widehat{\sigma_u}$
64×64	99.6094000000	0.0975000000	99.6092433089	0.0989692547	33.4635416667	0.3697318566	33.4462493563	0.3741631181
128×128	99.6094000000	0.0487000000	99.6097590990	0.0486867022	33.4635416667	0.1848659283	33.4537322188	0.1858105271
256×256	99.6094000000	0.0244000000	99.6096636839	0.0244907014	33.4635416667	0.0924329642	33.4595629123	0.0919732060
512×512	99.6094000000	0.0122000000	99.6095651442	0.0121198368	33.4635416667	0.0462164821	33.4654786002	0.0453526000
1024×1024	99.6094000000	0.0061000000	99.6096801758	0.0061338739	33.4635416667	0.0231082410	33.4640661364	0.0231559551

TABLE IV. NPCR RANDOMNESS TEST FOR IMAGE ENCRYPTION

Tested Image Size M-by-N 256-by-256		Theoretically NPCR Critical Value $\mathcal{N}^*_{0.05}$ =99.5693%	$\mathcal{N}^*_{0.01}$ =99.5527%	$\mathcal{N}^*_{0.001}$=99.5341%
Image Encryption Methods	Reported Value(s)	NPCR Test Results 0.05-level	0.01-level	0.001-level
Zhang 2005 [7]	98.669%	Fail	Fail	Fail
Zhu 2006 [8] (reported in [9])	99.26%	Fail	Fail	Fail
	99.45%	Fail	Fail	Fail
	99.13%	Fail	Fail	Fail
Behnia 2008 [6]	41.962%	Fail	Fail	Fail
Huang 2009 [9]	99.42%	Fail	Fail	Fail
	99.54%	Fail	Fail	Pass
	99.60%	Pass	Pass	Pass
Liao 2010 [10]	99.66%	Pass	Pass	Pass
	99.65%	Pass	Pass	Pass
	99.63%	Pass	Pass	Pass
Zhang 2010 [11]	99.61%	Pass	Pass	Pass
Kumar 2011 [12]	99.72%	Pass	Pass	Pass

Tested Image Size M-by-N 512-by-512		Theoretically NPCR Critical Value $\mathcal{N}^*_{0.05}$ =99.5893%	$\mathcal{N}^*_{0.01}$ =99.5810%	$\mathcal{N}^*_{0.001}$=99.5717%
Image Encryption Methods	Reported Value(s)	NPCR Test Results 0.05-level	0.01-level	0.001-level
Chen 2004 [5]	50.22%	Fail	Fail	Fail
Lian 2005 [13] (reported in [14])	99.5914%	Pass	Pass	Pass
Zhu 2010 [14]	99.6273041%	Pass	Pass	Pass

TABLE V. UACI RANDOMNESS TEST FOR IMAGE ENCRYPTION

Tested Image Size M-by-N 256-by-256		Theoretically UACI Critical Values $\mathcal{U}^{*-}_{0.05}$ =33.2824% $\mathcal{U}^{*+}_{0.05}$ =33.6447%	$\mathcal{U}^{*-}_{0.01}$ =33.2255% $\mathcal{U}^{*+}_{0.01}$ =33.7016%	$\mathcal{U}^{*-}_{0.001}$=33.1594% $\mathcal{U}^{*+}_{0.001}$ =33.7677%
Image Encryption Methods	Reported Value(s)	NPCR Test Results 0.05-level	0.01-level	0.001-level
Zhang 2005 [7]	33.362%	Pass	Pass	Pass
Zhu 2006 [8] (reported in [9])	21.41%	Fail	Fail	Fail
	23.42%	Fail	Fail	Fail
	15.08%	Fail	Fail	Fail
Behnia 2008 [6]	33.25%	Fail	Pass	Pass
Huang 2009 [9]	27.78%	Fail	Fail	Fail
	27.66%	Fail	Fail	Fail
	24.94%	Fail	Fail	Fail
Liao 2010 [10]	33.20%	Fail	Fail	Pass
	33.31%	Pass	Pass	Pass
	34.61%	Fail	Fail	Fail
Zhang 2010 [11]	38%	Fail	Fail	Fail
Kumar 2011 [12]	32.821%	Fail	Fail	Fail

Tested Image Size M-by-N 512-by-512		Theoretically UACI Critical Values $\mathcal{U}^{*-}_{0.05}$ =33.3730% $\mathcal{U}^{*+}_{0.05}$ =33.5541%	$\mathcal{U}^{*-}_{0.01}$ =33.3445% $\mathcal{U}^{*+}_{0.01}$ =33.5826%	$\mathcal{U}^{*-}_{0.001}$=33.3115% $\mathcal{U}^{*+}_{0.001}$ =33.6156%
Image Encryption Methods	Reported Value(s)	NPCR Test Results 0.05-level	0.01-level	0.001-level
Chen 2004 [5]	25.21%	Fail	Fail	Fail
Lian 2005 [13] (reported in [14])	33.3359%	Pass	Pass	Pass
Zhu 2010 [14]	33.4815979%	Pass	Pass	Pass

IV. Simulation Results

In this section, two types of simulations are presented. First, the Monte Carlo simulation is applied to generate interested statistics of $\mathcal{N}(I^1, I^2)$ and $\mathcal{U}(I^1, I^2)$, where I^1 and I^2 are images of size M-by-N generated by pseudo random number generator which is built-in function in MATLAB. Secondly, the designed NPCR and UACI tests are applied to various existing image encryption methods/ciphers.

A. Monte Carlo Simulation

In order to estimate the interested statistics, the sample mean and variance defined in Eqns. (21) and (22) are used, where X denotes the interested statistics and K is the number of observations. Recall the Law of Large Numbers (LLN), which states that the sample mean $\bar{X}$ converges to the true mean μ_X, as $K \to \infty$. Meanwhile, the sample variance is an unbiased and consistent estimator of the true variance, which implies that $S^2 \to \sigma_X^2$, as $K \to \infty$. Therefore, these two quantities can be used to estimate our interested statistics, including $\mu_{\mathcal{N}}$, $\sigma_{\mathcal{N}}$, $\mu_{\mathcal{U}}$ and $\sigma_{\mathcal{U}}$ under different F values.

$$\bar{X} = \sum_{i=1}^{K} X_i / K \tag{21}$$

$$S^2 = \sum_{i=1}^{K} (X_i - \bar{X})^2 / (K-1) \tag{22}$$

Simulation results of these interested statistics are shown in Table III. It is worth to note that each estimated statistics in Table III (marked with a cap), it is calculated from 10,000 pairs of I^1 and I^2 that are randomly generated images. More specifically, the estimated statistics $\widehat{\mu_{\mathcal{N}}}$, $\widehat{\sigma_{\mathcal{N}}}$, $\widehat{\mu_{\mathcal{U}}}$ and $\widehat{\sigma_{\mathcal{U}}}$ are obtained via Eqns. (23) –(26), respectively.

$$\widehat{\mu_{\mathcal{N}}} = \sum_{i=1}^{10000} \mathcal{N}(I_i^1, I_i^2) / 10000 \tag{23}$$

$$\widehat{\sigma_{\mathcal{N}}}^2 = \sum_{i=1}^{10000} (\mathcal{N}(I_i^1, I_i^2) - \widehat{\mu_{\mathcal{N}}})^2 / 9999 \tag{24}$$

$$\widehat{\mu_{\mathcal{U}}} = \sum_{i=1}^{10000} \mathcal{U}(I_i^1, I_i^2) / 10000 \tag{25}$$

$$\widehat{\sigma_{\mathcal{U}}}^2 = \sum_{i=1}^{10000} (\mathcal{U}(I_i^1, I_i^2) - \widehat{\mu_{\mathcal{U}}})^2 / 9999 \tag{26}$$

Fig. 2 shows the difference between the theoretical values and the experimental values. It is noticeable that such differences are subtle. More specifically, they are of or below the level of 10^{-4}. Therefore, the provided reference Tables I and II are reliable.

B. Randomness Test for Image Encryption

In this section, the reported results of differential attacks from various image encryption papers are collected and compared with critical values of NPCR and UACI tests.

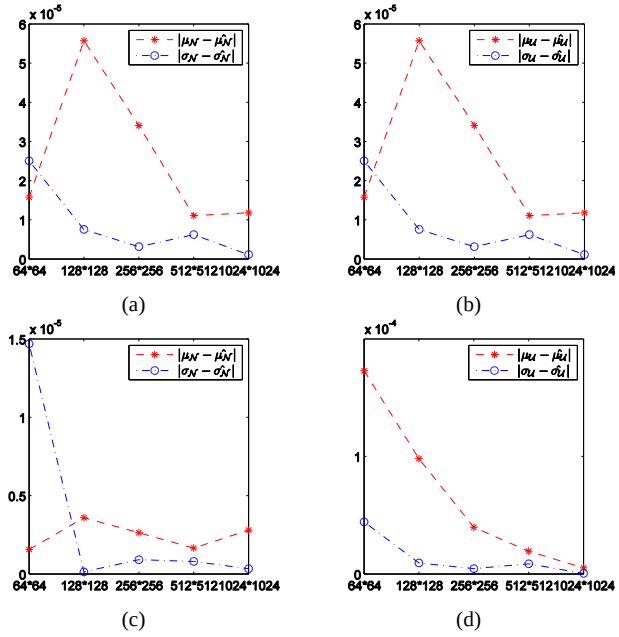


Fig. 2. Difference between the estimated values and experimental values (a) $|\mu_{\mathcal{N}} - \widehat{\mu_{\mathcal{N}}}|$ and $|\sigma_{\mathcal{N}} - \widehat{\sigma_{\mathcal{N}}}|$ when $F = 1$; (b) $|\mu_{\mathcal{U}} - \widehat{\mu_{\mathcal{U}}}|$ and $|\sigma_{\mathcal{U}} - \widehat{\sigma_{\mathcal{U}}}|$ when $F = 1$; (c) $|\mu_{\mathcal{N}} - \widehat{\mu_{\mathcal{N}}}|$ and $|\sigma_{\mathcal{N}} - \widehat{\sigma_{\mathcal{N}}}|$ when $F = 255$; (d) $|\mu_{\mathcal{U}} - \widehat{\mu_{\mathcal{U}}}|$ and $|\sigma_{\mathcal{U}} - \widehat{\sigma_{\mathcal{U}}}|$ when $F = 255$.

These image encryption methods include Zhang's method based on chaotic maps (Zhang 2005) [7], Zhu's method based on Chen's chaotic system (Zhu 2006) [8], Huang's method using multiple chaotic systems (Huang 2009) [9], Behnia's method using a mixture of chaotic maps (Behnia 2008) [6], Liao's algorithm based on self-adaptive wave transmission (Liao 2010) [10], Zhang's method using DNA addition with chaotic maps (Zhang 2010) [11], Kumar's method using extended substitution-diffusion network with chaos (Kumar 2011) [12], Chen's encryption scheme using the 3D cat map (Chen 2004) [5], Lian's block cipher using chaotic standard map (Lian 2005) [13], and Zhu's method using a bit-level permutation (Zhu 2010) [14]. The NPCR and UACI scores are obtained directly from papers of related methods without any modification.

Using reference Table I and II, these reported NPCR and UACI scores are evaluated to see whether the two test ciphertext images are random-like. In order to simplify the comparison, we listed these results in the chronological order and sorted with respect to the test images size, which determines the critical value(s) of the test. The NPCR and UACI test results are shown in Table IV and Table V, respectively.

From Table IV, it is noticeable that when the test image size is 256-by-256, although most NPCR scores are not far different from each other and close to 100%, they do have significant difference in the point view of statistics. Many earlier methods (before 2010) fail the test, but recent methods have better NPCR test results. Same phenomenon is also observed when the test image size is 512-by-512.

From Table V, it is clear that most of the test image encryption methods fail the UACI test, with an either too low or

too high UACI score.

Considering these results in Table IV and Table V, 'Lian 2005' [13] and 'Zhu 2010' [14] are two best ones among the test ten image encryption algorithms, because they passed the both the NPCR and UACI randomness tests. Although 'Zhu 2010' has slightly higher NPCR and UACI scores than those of 'Lian 2005', it does not mean that 'Zhu 2010' is more secure than 'Lian 2005', because their test scores are not statistically different. This conclusion also points out a common mistake in the image encryption literature: some author claims his/her method is better than some others' by simply comparing some test scores. For example, 'Lian 2005' [13] is used as a reference algorithm for comparing the NPCR and UACI scores with 'Zhu 2010' in [14], where the author claims that 'Zhu 2010' is better than 'Lian 2005' by simply comparing test scores. However, test results of 'Zhu 2010' and 'Lian 2005' in Tables IV and V show that they do not have significant difference. This implies that maybe both algorithms are able to generate random-like ciphertext image and thus the different test scores are purely caused by the stochastic process.

V. Conclusion

In this paper, we discussed the NPCR and UACI randomness tests for image encryption. Unlike the conventional usage of NPCR and UACI for calculating scores, we consider both scores as random variables under the ideally encrypted image model and derive their expectations and variances. Meanwhile, hypothesis tests with an α-level of significance are designed for NPCR and UACI tests respectively. With these two hypothesis tests, it is easy to accept or reject the null hypothesis that test ciphertext images are random-like. Therefore, such tests provide qualitatively results rather than quantitatively results for image encryption.

Experimental results show the estimated expectations and variance of NPCR and UACI are very close to the theoretical values, which justify the validity of theoretical values. Further, the proposed NPCR and UACI randomness tests are also applied to various image encryption algorithms. Test results show that many of these tested algorithms are problematic or at least not statistically random-like. Meanwhile, these results also showed that the conventionally quantitative analysis methodology for image encryption is questionable. Because these test scores, e.g. NPCR or UACI, are random variables dependent on parameters such as the image size and the format of the image rather than static values. Purely comparing two NPCR/UACI scores for two algorithms without noting these parameters is not fair. For example, a NPCR score based on gray images is 99.5710%, which is very close to the expectation 99.6094% (see Table I), but when the test image size is 512-by-512, this score is out of 99.9% confidence interval (99.5717%, 100%] of the NPCR score. This conclusion means that test ciphertext images do not follow the relations between two ideally encrypted images and thus it may be vulnerable to differential attacks. Moreover, the significance level α tells that the chance of making a wrong conclusion is one out of a thousand.

On the other hand, judging two encryption methods by comparing their test scores quantitatively is also questionable. In other words, better than some poor method(s)/algorithm(s) is not sufficient to say a method is good. Because it is still unclear whether this method is able to generate ciphertext images as random-like as those ideally encrypted images, although its test score is better than some other(s). Unless comparing test score(s) with theoretical values like those derived in this paper, it is hard to know whether a method is good and how good it is.

References

[1] E. Biham and A. Shamir, "Differential Cryptanalysis of DES-like Cryptosystems," in *Proceedings of the 10th Annual International Cryptology Conference on Advances in Cryptology*: Springer-Verlag, 1991.

[2] E. Biham and A. Shamir, "Differential Cryptanalysis of the Full 16-Round DES," in *Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology*: Springer-Verlag, 1993.

[3] "FIPS PUB 46: Data Encryption Standard," National Bureau of Standards, 1977.

[4] H. Williams, A. Webster, and S. Tavares, "On the Design of S-Boxes," in *Advances in Cryptology — CRYPTO '85 Proceedings*. vol. 218: Springer Berlin / Heidelberg, 1986, pp. 523-534.

[5] G. Chen, Y. Mao, and C. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," *Chaos, Solitons and Fractals,* vol. 21, pp. 749-761, 2004.

[6] S. Behnia, A. Akhshani, H. Mahmodi, and A. Akhavan, "A novel algorithm for image encryption based on mixture of chaotic maps," *Chaos, Solitons and Fractals,* vol. 35, pp. 408-419, 2008.

[7] L. Zhang, X. Liao, and X. Wang, "An image encryption approach based on chaotic maps," *Chaos, Solitons & Fractals,* vol. 24, pp. 759-765, 2005.

[8] C. X. Zhu, Z. G. Chen, and W. W. Ouyang, "A new image encryption algorithm based on general Chen's chaotic system," *Journal of Central South University (Science and Technology),* 2006.

[9] C. K. Huang and H. H. Nien, "Multi chaotic systems based pixel shuffle for image encryption," *Optics Communications,* vol. 282, pp. 2123-2127, 2009.

[10] X. Liao, S. Lai, and Q. Zhou, "A novel image encryption algorithm based on self-adaptive wave transmission," *Signal Processing,* vol. 90, pp. 2714-2722, 2010.

[11] Q. Zhang, L. Guo, and X. Wei, "Image encryption using DNA addition combining with chaotic maps," *Mathematical and Computer Modelling,* vol. 52, pp. 2028-2035, 2010.

[12] A. Kumar and M. K. Ghose, "Extended substitution-diffusion based image cipher using chaotic standard map," *Communications in Nonlinear Science and Numerical Simulation,* vol. 16, pp. 372-382, 2011.

[13] S. Lian, J. Sun, and Z. Wang, "A block cipher based on a suitable use of the chaotic standard map," *Chaos, Solitons & Fractals,* vol. 26, pp. 117-129, 2005.

[14] Z.-l. Zhu, W. Zhang, K.-w. Wong, and H. Yu, "A chaos-based symmetric image encryption scheme using a bit-level permutation," *Information Sciences,* vol. 181, pp. 1171-1186, 2010.

[15] M. Yang, N. Bourbakis, and L. Shujun, "Data-image-video encryption," *Potentials, IEEE,* vol. 23, pp. 28-34, 2004.

[16] "FIPS PUB140-1: Security Requirements for Cryptographic Modules." vol. 11: National Institute of Standards and Technology, 1994.

[17] "FIPS PUB140-2: Security Requirements for Cryptographic Modules," National Institute of Standards and Technology, 2001.

[18] Y. Mao, G. Chen, and S. Lian, "A novel fast image encryption scheme based on 3D chaotic Baker maps," *Int. J. Bifurcation and Chaos in June,* 2003.

[19] R. Peck, C. Olsen, and J. L. Devore, *Introduction to Statistics and Data Analysis*: Cengage Learning, 2008.

[20] M. Sternstein, *Statistics*: Barron's Educational Series, 1996.

[21] R. J. Larsen and M. L. Marx, *An introduction to mathematical statistics and its applications*: Pearson Prentice Hall, 2006.