J. Korean Math. Soc. **50** (2013), No. 4, pp. 879–897
http://dx.doi.org/10.4134/JKMS.2013.50.4.879

# ONLINE TEST BASED ON MUTUAL INFORMATION FOR TRUE RANDOM NUMBER GENERATORS

Young-Sik Kim, Yongjin Yeom, and Hee Bong Choi

Abstract. Shannon entropy is one of the widely used randomness measures especially for cryptographic applications. However, the conventional entropy tests are less sensitive to the inter-bit dependency in random samples. In this paper, we propose new online randomness test schemes for true random number generators (TRNGs) based on the mutual information between consecutive $k$-bit output blocks for testing of inter-bit dependency in random samples. By estimating the block entropies of distinct lengths at the same time, it is possible to measure the mutual information, which is closely related to the amount of the statistical dependency between two consecutive data blocks. In addition, we propose a new estimation method for entropies, which accumulates intermediate values of the number of frequencies. The proposed method can estimate entropy with less samples than Maurer-Coron type entropy test can. By numerical simulations, it is shown that the new proposed scheme can be used as a reliable online entropy estimator for TRNGs used by cryptographic modules.

## 1. Introduction

In cryptographic applications, random number generators are used to generate session keys, nonces, and prime numbers for digital signature and public key cryptography such as RSA and elliptic curve cryptography (ECC). If generated random numbers can be predictable by using the previous and/or the next values or they are not statistically independent from each other, the attacker can significantly reduce the complexity of the brute force attack. Consequently, the entire security of the crypto system can be vulnerable due to the weakness of the used random numbers.

In general, random number generators can be categorized into pseudo random number generators (PRNGs) and true random number generators (TRNGs). For the case of PRNGs, the initial values should be regularly obtained from the output of a TRNG in order to generate unpredictable random numbers.

Received September 19, 2012.

2010 *Mathematics Subject Classification.* Primary 94A60; Secondary 11K45.

*Key words and phrases.* online test, TRNG, random number generation, statistical test, Shannon entropy, mutual information.

That is, the unpredictability of a PRNG is solely dependent on the entropy of the random seed generated by a TRNG. Therefore, commercial smart cards include a TRNG for their security and the random seed generated from the TRNG should be used for initializing PRNG [9].

TRNGs are random number generators based on physical noise sources [1, 10] so as to produce unpredictable random numbers, while they can be easily influenced by the environmental causes. For example, it is possible to attack TRNGs based on the thermal noise by controlling the background temperature [1]. Moreover, due to the aging effect, the statistical quality of the TRNG will be deteriorated. Therefore, the security standard for the TRNG such as the German standard AIS.31 requires that the TRNG should equip not only a post-processing method which can fix some statistical bias in the TRNG output, but also an online test method to check the statistical quality of the TRNG output on the fly [5].

Shannon entropy is used as a measure of randomness. However, widely used Shannon entropy estimation methods are less sensitive to the variation of dependency between consecutive random data. If there is a statistical dependency between random data, the adversary exploits this property to predict more accurately what comes next. Therefore, we need a method to measure the statistical dependency as well as the uniformness.

In this paper, we propose new online test schemes for TRNG which evaluate the mutual information between consecutive $k$-bit random output blocks. That is, the mutual information will be used as a measure of dependency between consecutive bits. For a stationary random source, it will be shown that the mutual information can be estimated by using block entropies with distinct sizes. In addition, it will be shown that by estimating the block entropies with distinct sizes simultaneously, we can measure efficiently the mutual information of random data which are generated from a stationary random source. Moreover, the proposed schemes can be easily implemented by using a small amount of additional memory.

## 2. Preliminaries

### 2.1. Statistical tests for randomness

It is not an easy task to decide whether a given sequence is random or not. Especially, it is very difficult to verify the randomness of a TRNG output based on physical behaviors of the device. Therefore, for practical applications, the decision might be based on a series of the statistical tests for the TRNG output samples with a finite length.

The statistical quality of random sequences can be tested based on two categories: the uniformness and the statistical independence. While the uniformness generally can be measured using relatively simple methods such as counting the numbers of occurrences of each symbol, evaluating the statistical independence is a relatively difficult task to implement and usually checked by several indirect tests on random sequences [5]. Moreover, it is difficult to implement those statistical tests at the same time for an online test which runs for cryptographic modules in embedded devices because of the limited resources in systems.

The amount of the tolerance level for the statistical weakness is dependent on applications. This tolerance level can be defined as the lower or the upper bounds of the probability of occurrence of a noise alarm in the course of one year typical use of the TRNG. For example, in the German standard AIS.31, the lower bound of the above probability is specified as $10^{-6}$ [5]. According to a given tolerance level, the threshold of the test as well as the probability of detection can be determined.

One of the widely used online test schemes is a test based on the $\chi^2$ test with degree of freedom of 15 for 128 4-bit blocks [7]. In this test, the frequencies of 16 patterns representing a 4-bit block are counted and the counted values are subtracted by the mean value, squared, and averaged by all values from the 16 patterns. This method is simple but cannot detect all kinds of statistical defects.

### 2.2. Entropy as a randomness measure

C. E. Shannon introduced the concept of entropy in order to measure the quantity of information represented by a discrete random variable $X$ as follows [4]

$$H(X) = -\sum_{x} \Pr(x) \log_2 \Pr(x).$$

The entropy $H(X)$ is usually interpreted as the uncertainty of the random variable $X$. This is the most widely used definition of the entropy in various areas including communication and security. Especially in security applications, Shannon entropy is used as a randomness measure for a random source [6, 2, 3].

It is assumed that the random source is stationary. That is, we assume that statistical characteristics of the random source are independent of observation interval. The random source generates continuous random data. We denote by *random data* the output of TRNG which can be tested. By splitting the serial random data into disjoint but fixed length bit blocks, we can estimate the block Shannon entropies for the fixed sample space whose elements are $2^k$ $k$-bit data blocks. For some proposals [6, 3], the block entropy is used instead of the normal Shannon entropies. In this paper, we will confine our interest to block Shannon entropy, shortly, block entropy. Note that for the block size $k$, the maximum entropy value is $k$ when every block is equiprobable. For example, when the block size $k$ equals 8 as in Figs. 5 and 6 in Section 4, the maximum entropy value is 8.

If a random variable $Y$ is given, then the entropy of the other random variable $X$ is represented as a conditional entropy defined as [4]

$$\begin{aligned} H(X|Y) &= -\sum_{x,y} \Pr(x,y)\log_2 \Pr(x|y) \\ &= H(X,Y) - H(Y). \end{aligned} \tag{1}$$

The conditional entropy $X$ given $Y$ is interpreted as the remaining uncertainty of the random variable $X$ when the random variable $Y$ is disclosed. If two random variables are correlated, then disclosure of one random variable reduces the uncertainty of the other random variable. The amount of the shared uncertainty of two random variables is measured using the mutual information defined as

$$I(X;Y) = \sum_{x,y} \Pr(x,y)\log_2 \frac{\Pr(x,y)}{\Pr(x)\Pr(y)}.$$

### 2.3. Weakness of the conventional entropy estimation methods

The main problem of the widely used Shannon entropy estimation methods as a randomness measure is that it is less sensitive to the variation of the statistical dependency between consecutive bits. For example, if a sequence with heavy bias is given, the estimated Shannon entropy tends to be low as expected. However, if a sequence with significant statistical dependencies between neighbor bits is given, the estimated Shannon entropy of the sequence is relatively high.

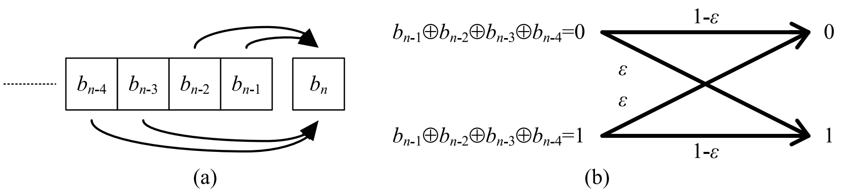
$b_{n\text{-}4}$
$b_{n\text{-}3}$
$b_{n\text{-}2}$
$b_{n\text{-}1}$
$b_n$
(a)
$b_{n\text{-}1}\oplus b_{n\text{-}2}\oplus b_{n\text{-}3}\oplus b_{n\text{-}4}=0$
$b_{n\text{-}1}\oplus b_{n\text{-}2}\oplus b_{n\text{-}3}\oplus b_{n\text{-}4}=1$
1-ε
ε
ε
1-ε
0
1
(b)

FIGURE 1. Dependency model.

In order to test the strength of the randomness testing method with respect to the statistical dependency, the dependency model depicted in Fig. 1 is considered. Fig. 1 shows that the $n$th bit is influenced by the previous four bits, $b_{n-4}$, $b_{n-3}$, $b_{n-2}$, and $b_{n-1}$. That is, if the exclusive OR sum of the 4 previous bits are 0, i.e., $b_{n-1} \oplus b_{n-2} \oplus b_{n-3} \oplus b_{n-4} = 0$, then the probability of $b_n = 0$ is $1 - \epsilon$ and the probability of $b_n = 1$ is $\epsilon$, where $0 < \epsilon < 0.5$. Otherwise, if $b_{n-1} \oplus b_{n-2} \oplus b_{n-3} \oplus b_{n-4} = 1$, then the probability of $b_n = 0$ is $\epsilon$ and the opposite probability is $1 - \epsilon$. The random test vectors are generated according to the dependency model in Fig. 1. It is easy to generalize the dependency model in Fig. 1 for $k$-bit dependency. Especially, for 1-bit dependency, the dependency model corresponds to the case of $\Pr(b_n = 1|b_{n-1} = 1) = \Pr(b_n = 0|b_{n-1} = 0) = 1 - \epsilon$, which is a straightforward way to present the statistical dependency between two consecutive bits.

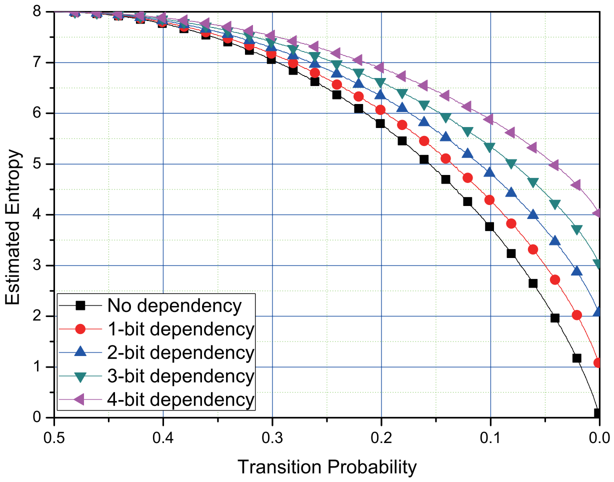

FIGURE 2. Shannon entropy as a measure of statistical independence of the random numbers.

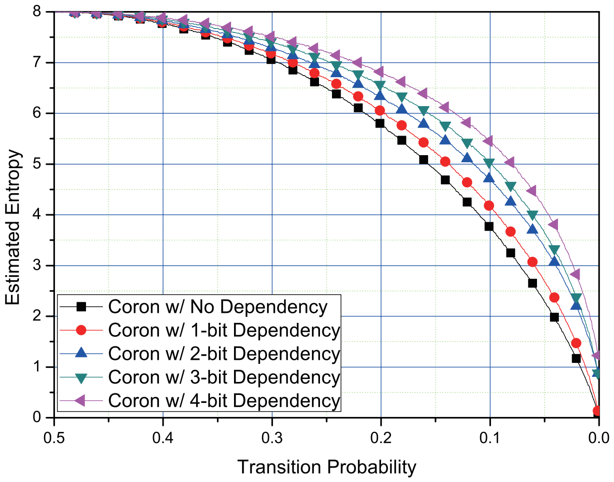

FIGURE 3. Coron's entropy test as a measure of statistical independence of the random numbers.

Then the weakness of the entropy estimation methods is illustrated in Figs. 2 and 3, which show the entropy estimation results by using the frequency counting method presented in Subsection 3.3.1 and Coron's entropy test presented in Subsection 3.3.2, respectively. In Figs. 2 and 3, the horizontal axis corresponds to the transition probability $\epsilon$, which varies between 0 and 0.5. For the case of the frequency counting method in Fig. 2, the entropies of the data with $k$-bit dependency ($k = 1, 2, 3, 4$) are higher than that of random data. For the case of Coron's entropy test in Fig. 3, the estimated entropy is still higher than the expected value given the bias even though it is more sensitive to the variation of the transition probability than the frequency counting method. Therefore, we can think that there is a difficulty to distinguish the difference between relatively high entropy values and higher dependency between subsequent bits when the widely used Shannon entropy estimation methods are applied as a randomness measure. In other words, although the given sequence violates not only the uniformity but also the statistical independence, the estimated Shannon entropy of the random sequence may indicate that the given sequence is sufficiently random.

We can understand the result in Fig. 2 especially for one bit dependency given as $\Pr(b_i = 0|b_{i-1} = 0) = \Pr(b_i = 1|b_{i-1} = 1) = 1 - \epsilon$, $\Pr(b_i = 1|b_{i-1} = 0) = \Pr(b_i = 0|b_{i-1} = 1) = \epsilon$. Therefore, for very small $\epsilon$, 0 or 1 will be continuously generated for a long time with high probability. However, with $1/\epsilon$ bits for $\epsilon > 0$, we can expect the bit transition 0 to 1 or 1 to 0. Therefore, two kinds of patterns (all ones and all zeros) alternatively appear in long enough output. If we split the output into $k$-bit blocks, 00...0 and 11...1 will be the dominant patterns among $2^k$ block patterns within the whole sequence. Therefore, Shannon entropy is expected to be over 1 in this case. We can similarly expand this explanation to the multi-bit dependency.

Therefore, we need a new method to check the statistical dependency in random samples. Especially, since TRNGs are based on the physical random source, the statistical dependency is an important factor for the quality of the generated random data. For example, consider the ring-oscillator (RO) based TRNGs which use several ROs to increase the throughput [8]. If a set of ROs are closely located in a chip due to the restricted hardware resources, then they can be electro-magnetically coupled and as a result, the generated random data can be statistically dependent to their neighbor bits. In the next section, we will propose a new testing method based on the mutual information.

## 3. Mutual information based online test

### 3.1. Theoretical backgrounds

Suppose that $n$-bit data $b_1, b_2, \ldots, b_n$ are generated from a TRNG. The binary sequence is divided into several $k$-bit blocks of the form

$$(b_1, \ldots, b_k),(b_{k+1}, \ldots, b_{2k}), \ldots = \mathbf{b}_1, \mathbf{b}_2, \ldots,$$

where $\mathbf{b}_i = (b_{(i-1)k+1}, \ldots, b_{ik})$ for $i = 1, 2, \ldots$. The proposed online test carries out a random test based on each block $\mathbf{b}_i$ for checking the statistical dependency between two adjacent blocks $\mathbf{b}_j$ and $\mathbf{b}_{j+1}$ for all $j$.

Note that there are $2^k$ possible patterns for each block $\mathbf{b}_i$. Let $B_i$ be the random variable of $i$th random block. If the two consecutive blocks are statistically independent from each other and are stationary, then the following property is satisfied. The relation between the joint probability of $B_i = \mathbf{b}_i$ and $B_{i+1} = \mathbf{b}_{i+1}$ and the marginal probabilities is given as

$$\begin{aligned}\Pr(B_i = \mathbf{b}_i, B_{i+1} = \mathbf{b}_{i+1}) &= \Pr(B_i = \mathbf{b}_i) \cdot \Pr(B_{i+1} = \mathbf{b}_{i+1}) \\ &= [\Pr(B_i = \mathbf{b}_i)]^2.\end{aligned}$$

Then we can check a relation between the mutual information and block entropies as in the following proposition.

**Proposition 3.1.** *For a stationary random source, the mutual information between two consecutive $k$-bit random blocks can be calculated as*

$$I(B_{i+1}; B_i) = 2H(B_i) - H(B_{i+1}B_i). \tag{2}$$

*Proof.* For testing statistical dependency between blocks, we can measure the mutual information between two blocks as $I(B_i; B_{i+1})$, which can be rewritten as

$$I(B_{i+1}; B_i) = H(B_i) + H(B_{i+1}) - H(B_{i+1}B_i). \tag{3}$$

For the stationary random source, we have $H(B_i) = H(B_{i+1})$. Finally, we have $I(B_{i+1}; B_i) = 2H(B_i) - H(B_{i+1}B_i)$. □

### 3.2. Generalization of the mutual information for online test

In this subsection, we will generalize Proposition 3.1 to the mutual information estimation with longer message random variables in order to measure long-term dependency of the produced random data blocks.

**Proposition 3.2.** *The mutual information $I(B_{i+n-1} \cdots B_{i+1}B_i; B_{i+n})$ can be evaluated by estimating the following three entropies for distinct blocks of sizes $nk$, $k$, and $(n+1)k$*

$$\begin{aligned}I(B_{i+n-1} \cdots B_{i+1}B_i; B_{i+n}) = {} & H(B_{i+n-1} \cdots B_{i+1}B_i) + H(B_{i+n}) \\ & - H(B_{i+n} \cdots B_{i+1}B_i).\end{aligned} \tag{4}$$

*Proof.* Since $I(B_{i+n-1} \cdots B_{i+1}B_i; B_{i+n})$ is just a mutual information between two random variables with distinct lengths, we can similarly check the equality (4) as in Proposition 3.1. □

**Example 1.** Suppose that we are going to estimate $I(B_{i+1}B_i; B_{i+2})$. Then we have to evaluate three entropies $H(B_{i+1}B_i)$, $H(B_{i+2})$, and $H(B_{i+2}B_{i+1}B_i)$. From these values, the conditional mutual information can be calculated as

$$I(B_{i+1}B_i; B_{i+2}) = H(B_{i+1}B_i) + H(B_{i+2}) - H(B_{i+2}B_{i+1}B_i). \tag{5}$$

It is possible to evaluate the conditional mutual information as follows. Suppose that we are going to measure the mutual information between two blocks $B_{i+n-1}\cdots B_{i+1}$ and $B_{i+n}$ given $B_i$, i.e., $I(B_{i+n-1}\cdots B_{i+1}; B_{i+n}|B_i)$.

**Proposition 3.3.**

$$\begin{aligned} I(B_{i+n-1}\cdots B_{i+1}; B_{i+n}|B_i) = {} & H(B_{i+n-1}\cdots B_{i+1}B_i) + H(B_{i+n}B_i) \\ & - H(B_i) - H(B_{i+n-1}\cdots B_{i+1}B_{i+n}B_i). \end{aligned}$$

*Proof.* By the chain rule of the mutual information, we have

$$(6) \qquad I(B_{i+n-1}\cdots B_{i+1}; B_{i+n}|B_i) = \sum_{k=1}^{n} I(B_{i+k}; B_{i+n}|B_{i+k-1}\cdots B_i).$$

Using (1) and (3), the identity (6) can be rewritten as in the statement of this proposition. □

That is, the conditional mutual information $I(B_{i+n-1}\cdots B_{i+1}; B_{i+n}|B_i)$ can be evaluated by estimating four entropies for distinct block size.

**Example 2.** Suppose that we are going to estimate $I(B_{i+1}; B_{i+2} \mid B_i)$. Then we have to evaluate four entropies $H(B_{i+1}B_i)$, $H(B_{i+2}B_i)$, $H(B_i)$, and $H(B_{i+2}B_{i+1}B_i)$. From these values, the conditional mutual information can be calculated as

$$(7)$$
$$I(B_{i+1}; B_{i+2} \mid B_i) = H(B_{i+1}B_i) + H(B_{i+2}B_i) - H(B_i) - H(B_{i+2}B_{i+1}B_i).$$

For the stationary random source, we have $H(B_{i+1}B_i) = H(B_{i+2}B_i)$. Therefore, (7) can be simplified as

$$(8) \qquad I(B_{i+1}; B_{i+2} \mid B_i) = 2H(B_{i+1}B_i) - H(B_i) - H(B_{i+2}B_{i+1}B_i).$$

### 3.3. Design of online test

In estimating the joint Shannon entropies such as $H(B_iB_{i+1})$ and $H(B_i)$ in (2), (5), and (7), there are two known approaches: counting the frequency of each pattern or accumulating the minimum distance between the same patterns [6, 2, 3]. In this subsection, we briefly review these two known approaches and propose a new estimation method for Shannon entropy.

**3.3.1.** *Frequency counting.* This is a straightforward method of block entropy estimation. Suppose that there are $2^k$ memory buffers for each block pattern. These buffers are used to store the frequency of each pattern in the given random sequence. Suppose that $N$ blocks are counted for this test. Let $h(i)$ $(0 \le i < 2^k)$ be the counted values stored in each buffer. That is, the sum of all $h(i)$'s is equal to $N$ after finishing the counting. After finishing the counting of $N$ blocks, Shannon entropy is calculated as

$$f_{FC} = \sum_{i=0}^{2^k-1} \frac{h(i)}{N} \log_2 \frac{N}{h(i)}$$

$$= \log_2 N - \frac{1}{N} \sum_{i=0}^{2^k-1} h(i) \log_2 h(i). \tag{9}$$

**3.3.2.** *Accumulation of reappearance distance.* Maurer proposed a different approach to estimate Shannon entropy called as the universal statistical test [6]. Later, Coron refined Maurer's universal statistical test in order to remove the deviation from the real Shannon entropy values [3]. In their approaches, instead of storing the number of occurrences of each pattern, the indices of the patterns are stored in the $2^k$ buffers. At the same time, the difference value between the current pattern and the corresponding value stored in the memory is accumulated according to the following equation

$$\begin{aligned} f_{MC} &= \frac{1}{K} \sum_{n=Q+1}^{Q+K} g(A_n), \\ g(i) &= \frac{1}{\ln 2} \sum_{v=1}^{i-1} \frac{1}{v}, \end{aligned} \tag{10}$$

where $Q$ is the number of samples for initialization, $K$ is the number of samples for entropy evaluation, and $A_n$ is the difference between the index of the current $2k$-bit pattern and that of the stored value (previous index of the same pattern) in the corresponding buffer. Then it was shown that the test function $f_{MC}$ asymptotically converses to the entropy value of the given random sample [2].

**3.3.3.** *Accumulating intermediate frequencies*: *a new proposal.* In this subsection, we will propose another way to estimate Shannon entropy. In this method, for a given block size $k$, the frequency of each block ($2^k$ blocks) is counted as in the frequency counting method from Subsection 3.3.1. Each pattern is indexed by a numerical order as $i = \sum_{j=0}^{k-1} b_j 2^j$. Suppose that $h(\mathbf{b}_n = i)$ is the value of the $i$th counter when the $n$th block $\mathbf{b}_n$ corresponds to the pattern with index $i$. Hereafter, when the specific index number is not considered, we denote $h(\mathbf{b}_n = i)$ as $h(\mathbf{b}_n)$ by abuse of notation.

Because the initial contents of registers are set zero, we can introduce the initialization phase as for Coron's test in order to collect some rough statistics on the occurrence frequencies. After the initialization phase, the accumulation stage begins. In the accumulation stage, *the proposed test function* is evaluated according to the following way

$$f_{AF} = \frac{1}{K} \sum_{n=Q+1}^{Q+K} \log_2(n/h(B_n)).$$

That is, the proposed method is to take the time average of the logarithm of the intermediate frequency ratio of each pattern

$$f_{AF} = \frac{1}{K}\sum_{n=Q+1}^{Q+K} \log_2 n - \frac{1}{K}\sum_{n=Q+1}^{Q+K} \log_2[h(B_n)]. \tag{11}$$

Note the difference between (9) and (11). The test value (11) can be continuously updated during the counting.

The proposed test procedure can be summarized as in Algorithm 1.

**Algorithm 1.** Proposed algorithm for entropy estimation using accumulating intermediate frequencies.

```
1) Initialization Phase
    (1) Set $2^k$ memory blocks as zero.
    (2) While $n < Q$
        (a) Take $k$-bit data $\mathbf{b}_n = \{b_{kn}, b_{kn+1}, \ldots, b_{kn+k-1}\}$.
        (b) Increase $h(i)$ by 1 at the $i$th memory block when $i = \sum_{j=0}^{k-1} b_{kn+j}2^j$.
        (c) Increase $n$ by 1.
2) Evaluation Phase
    (1) Set $f_{AF} = 0$ and $C = 0$.
    (2) While $n < K + Q$
        (a) Take $k$-bit data $\mathbf{b}_n = \{b_{kn}, b_{kn+1}, \ldots, b_{kn+k-1}\}$.
        (b) Increase $C$ by 1.
        (c) Increase $h(i)$ by 1 at the $i$th memory block when $i = \sum_{j=0}^{k-1} b_{kn+j}2^j$.
        (d) $f_{AF} \leftarrow f_{AF} + \log_2 n/h(i)$.
        (e) Increase $n$ by 1.
    (3) $f_{AF} \leftarrow f_{AF}/C$.
```

It is easy to check that the expected value of the test function in (11) converges to the entropy of the stationary random source as in the following theorem.

**Theorem 3.4.** *The expected value of the test value $f_{AF}$ in* (11) *by Algorithm 1 is the entropy of random source.*

*Proof.* Suppose that $q_i$ is the probability of being $B_n = i$, where $i$ is the index of $n$th block. Then for large $n$, we have $q_i = q_i(n) = h(B_n = \mathbf{b}_n = i)/n$, where $i = \sum_{j=0}^{k-1} b_{kn+j}2^j$. Therefore, (11) can be rewritten as

$$\begin{aligned} f_{AF} &= \frac{1}{K}\sum_{n=Q+1}^{Q+K} \log_2 n - \frac{1}{K}\sum_{n=Q+1}^{Q+K} \log_2 nq_i \\ &= -\frac{1}{K}\sum_{n=Q+1}^{Q+K} \log_2 q_i. \end{aligned}$$

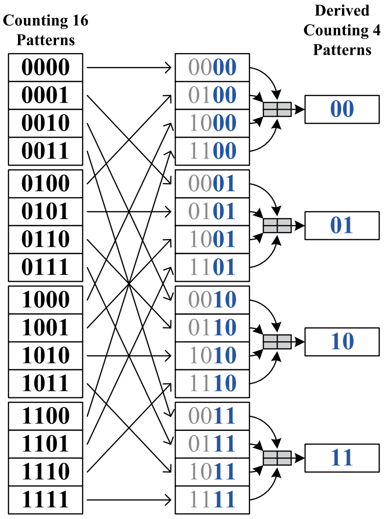

FIGURE 4. Derivation of frequencies of 2-bit patterns from the frequencies of 4-bit patterns.

For large $n$, $q_i$ converges to the probability of the pattern. Then the expectation of the summand $\log_2 q_i$ is given as

$$E[\log_2 h(B_n)] = -\sum_{i=0}^{2^k-1} q_i \log_2 q_i = H(B_n). \qquad \square$$

Therefore, for the ergodic source, we can obtain the entropy value from the proposed test functions.

**3.3.4.** *Comparison between entropy estimation tests.* In this subsection, we propose an efficient architecture for the proposed online test. For the estimation of the mutual information $I(B_i; B_{i+1})$ between consecutive blocks, we have to estimate both $H(B_iB_{i+1})$ and $H(B_i)$ at the same time, where it is necessary to use $2^{2k}$ and $2^k$ storage spaces. For the case of the frequency counting approaches in Subsections 3.3.1 and 3.3.3, it is possible to reduce the number of required storage space because specific sums of the number of occurrences of $2k$-bit patterns can be used as for that of $k$-bit patterns. Fig. 4 illustrates an example of the conversion from the number of 4-bit patterns to that of 2-bit patterns under the assumption of the stationary random source. That is, it is easy to see that if the upper $k$-bit can be treated as 'don't care' bit among $2k$-bit patterns, by summing of $2^k$ patterns with the same lower $k$ bits, we can obtain the number of occurrences of $2^k$ $k$-bit patterns from the number of occurrences of $2k$-bit patterns, which reduces $2^k$ storage spaces such as registers.

Table 1. Comparison of characteristics for each entropy estimation method

| | Freq. Counting | Maurer-Coron | New Proposition |
|---|---|---|---|
| Logarithmic table | Large | Small | Large |
| Number of Registers | $2^{2k}$ | $2^{2k} + 2^k$ | $2^{2k}$ |
| Samples | Fixed | Variable | Variable |
| Initialization | No | Yes | Yes |
| Convergence Speed | Fast | Slow | Fast |

Table 1 shows the comparison of major characteristics among three entropy estimation methods. In the first row, we see that all of them require logarithmic operations in (9), (10), and (11), which can be replaced with a lookup table. However, especially for the case of Coron's test, this table can be reduced to contain only the values of $\{\log_2 g(2), \log_2 g(3), \ldots, \log_2 g(L_{\max})\}$, where $L_{\max}$ is the maximum distance between two adjacent blocks with the same index. In the second row, as discussed earlier, for the cases of the frequency counting method and the proposed methods, the counting results for longer blocks such as $2k$-bit can be used to count the occurrences of shorter blocks such as $k$-bit. However, for the case of Coron's entropy test, it is not trivial to reduce the number of required storage space. In the third row, it is noted that the sample sizes for Coron's entropy test and the proposed scheme, which are based on the time average of updated information, is more flexible than the frequency counting method, which uses ensemble average. In the fourth row, we can see that the initialization phase is necessary for Coron's entropy test and the proposed scheme, but not for the frequency counting. In the last row, it is said that the frequency counting method and the proposed scheme require less number of samples to obtain a result with a specific accuracy than Coron's entropy test does, which will be illustrated by a numerical simulation in Subsection 4.2.

## 4. Numerical results

In this section, we present some numerical results for the new proposed entropy estimation method in Subsection 3.3.3 and the proposed online test scheme for TRNGs.

### 4.1. Simulation setting

For numerical simulations, the random sequences with the specified dependency based on the model in Fig. 1 are generated using a pseudo-random number generator. In order to generate random numbers with the specific transition probability $\epsilon$, firstly a pseudo-random number with a uniform distribution between $[0, 1]$ is generated. Then for the generation of binary random data, if this

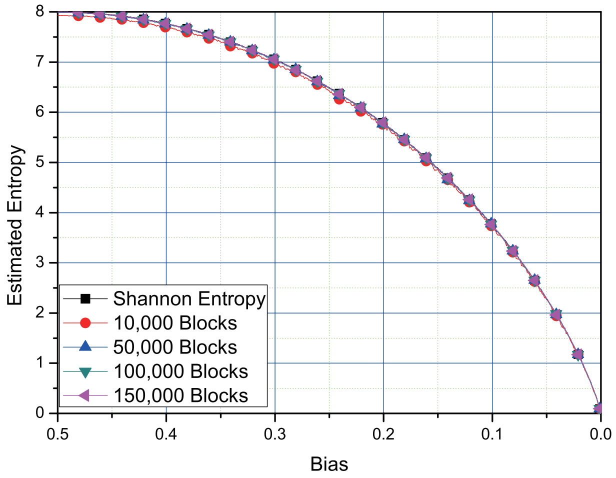

FIGURE 5. Performance of the proposed entropy estimation scheme in Subsection 3.3.3 for various sample sizes such as 10,000, 50,000, and 150,000 blocks.

random number is greater than the specified value $\epsilon$, then the next bit is decided as the inversion of exclusive OR (XOR) of the previous bits. Otherwise, the next bit is the same as XOR of the previous bits.

There are two reasons to adopt a pseudo randomly generated sequences to test the proposed scheme for TRNGs. First, it is highly difficult to generate a random sequence with the specified statistical characteristics directly from TRNGs. If we choose random sequences with specific statistical characteristics among millions of random samples from TRNGs, then it is hard to determine whether the collected random sequences actually have the characteristics or not because these characteristics are not intentionally generated, but experimentally estimated using some testing methods. In addition, a statistical test will work for both pseudo randomly and true randomly generated sequences. Therefore, it is reasonable to think that if a statistical test works for the pseudo randomly generated random sequences, then it will work for the true randomly generated random sequences also.

### 4.2. Performance of the proposed entropy estimation

As the first step, we check the performance of the proposed entropy estimation scheme presented in Subsection 3.3.3. Fig. 5 shows the estimation results of the proposed scheme with various sample sizes, 10,000, 50,000, 100,000, and 150,000 blocks. In this figure, the bias 0.5 means the perfectly balanced case, i.e., $\Pr(b_i = 0) = \Pr(b_i = 1) = 0.5$ for some $i$. We can see that as the sample size increases, the test value converges to block Shannon entropy with block

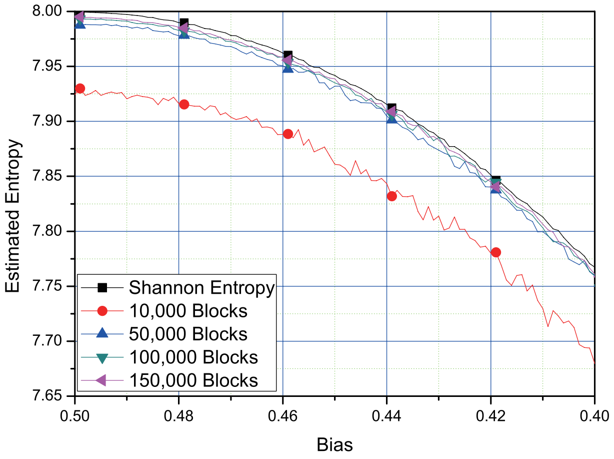

Figure 6. Performance of the proposed entropy estimation scheme in Subsection 3.3.3 for the bias between 0.4 to 0.5. This is an enlarged version of Fig. 5.

size $k = 8$. As we can see in Fig. 5, the test value is more deviated from the exact Shannon entropy around the perfectly balance case. Fig. 6 is the enlarged version of Fig. 5 around the bias from 0.4 to 0.5 in order to clearly show this deviation. Because for nearby the perfectly balanced case, every pattern occurs with almost the same probability. Therefore, samples for each pattern are relatively shorter than for highly biased case (i.e., $\epsilon \sim 0$), where few patterns are frequently occurred and enough samples are collected for those patterns.

Fig. 7 shows the effect of the initialization phase, which compares the simulation results of 10,000 blocks with no initialization, initialization of 256 blocks, 2,560 blocks, and 5,120 blocks. If the number of the initialization is increasing, the deviation is reducing.

Fig. 8 compares the proposed entropy estimation scheme with the Coron's entropy test with the same block size. Note that with the same number of blocks, the new proposition shows less fluctuations than the Coron's test does. That is, the new proposed scheme is more stable than the Coron's entropy test with less number of blocks. Also note that with less number of blocks, the results of the new proposed schemes are deviated from the actual Shannon entropy values. However, as the number of blocks is increased, the estimated value converges to the real Shannon entropy values for given bias.

### 4.3. Estimation of mutual information $I(B_{i+1}; B_i)$

Fig. 9 shows the estimation results for the random sequences which are intentionally generated to have 1-bit and 2-bit dependencies for the block size $k = 4$. The horizontal axis in Fig. 9 corresponds to the transition probability $\epsilon$, which

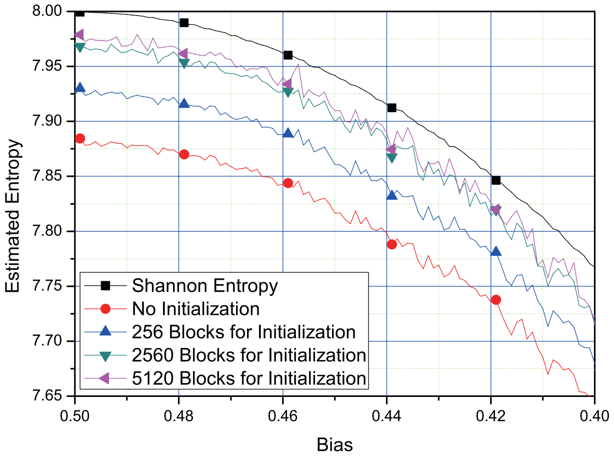

Figure 7. Performance of the proposed scheme in Subsection 3.3.3 for the distinct number of initialization blocks such as 0, 256, 2,560, and 5,120 blocks.

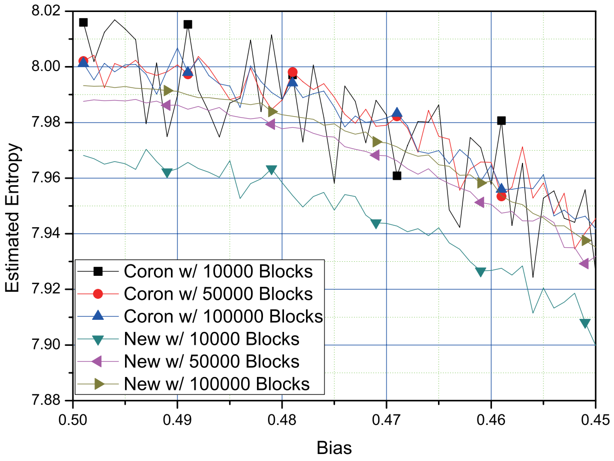

Figure 8. Comparison of Coron's entropy test in Subsection 3.3.2 and the proposed scheme in Subsection 3.3.3 for 10,000, 50,000 and 100,000 blocks.

can be presented as $\Pr(b_i \mid b_{i-1})$ for one-bit dependency and as $\Pr(b_i \mid b_{i-1}, b_{i-2})$ for two-bit dependency. Therefore, if the horizontal value approaches 0.5, we have $\Pr(b_i \mid b_{i-1}) \approx \Pr(b_i) \approx 0.5$ or $\Pr(b_i \mid b_{i-1}, b_{i-2}) \approx \Pr(b_i) \approx 0.5$, which means that the $i$th bit is close to being independent from the previous bits, *vise versa*. Because the block size is $k = 4$, we have estimation results for not

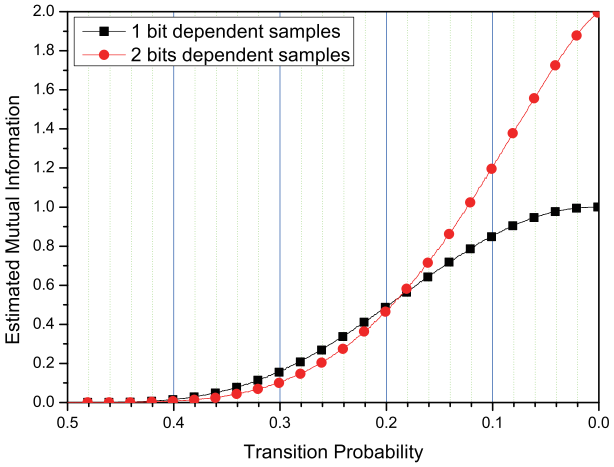

FIGURE 9. Estimated Mutual Information using Proposition 3.1 for $k = 4$ and 150,000 blocks.

only the mutual information between 4-bit blocks, but also for both 8-bit block entropy and 4-bit block entropy as in Fig. 10. Therefore, two thresholds can be set as one for the uniformness criterion, and the other for the independency criterion. For example, according to the AIS.31 [5], the threshold of the Coron's test with $k = 8$ is specified as 7.976. Similarly, we can experimentally set the independency threshold as 0.001, which is almost the same amount as the uniformness criterion. However, it is possible to set another threshold depending on the requirement of each application.

### 4.4. Estimation of mutual information $I(B_{i+2}B_{i+1}; B_i)$

Fig. 11 shows the estimation results for the mutual information $I(B_{i+2}B_{i+1}; B_i)$ for $k = 2$ with respect to the two bit dependency. Actually, this corresponds to the mutual information between the 4-bit block and the previous 2-bit block. In Fig. 11, the horizontal axis corresponds to the transition probability $\epsilon$. For the estimation of $I(B_{i+2}B_{i+1}; B_i)$, we have to estimate three values $H(B_{i+2}B_{i+1}B_i)$, $H(B_{i+1}B_i)$, and $H(B_i)$ whose maximum values are 6, 4, and 2, respectively. In result, three entropy values are greater than 2 even for very small $\epsilon$. Therefore, the mutual information is close to 2 at $\epsilon \to 0$.

### 4.5. Estimation of conditional mutual information $I(B_{i+2}; B_{i+1}|B_i)$

Fig. 12 shows the estimation results for the mutual information $I(B_{i+2}; B_{i+1}|B_i)$ for $k = 1$ with respect to the two bit dependency. Again, the horizontal axis in Fig. 12 corresponds to the transition probability $\epsilon$. For the estimation of $I(B_{i+2}B_{i+1}; B_i)$, we have to estimate three values $H(B_{i+2}B_{i+1}B_i)$, $H(B_{i+1}B_i)$, and $H(B_i)$ whose maximum values are 3, 2, and 1, respectively. In result, two

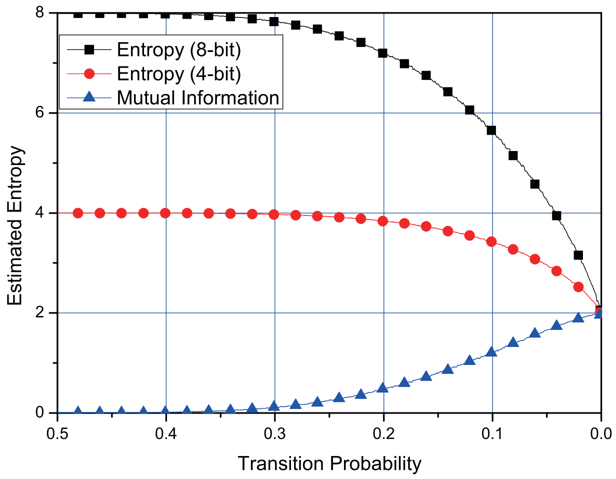

FIGURE 10. Estimated Mutual Information using Proposition 3.1 for $k = 4$ and 150,000 blocks. The estimated entropies for 8-bit/4-bit block sizes are depicted at the same time.

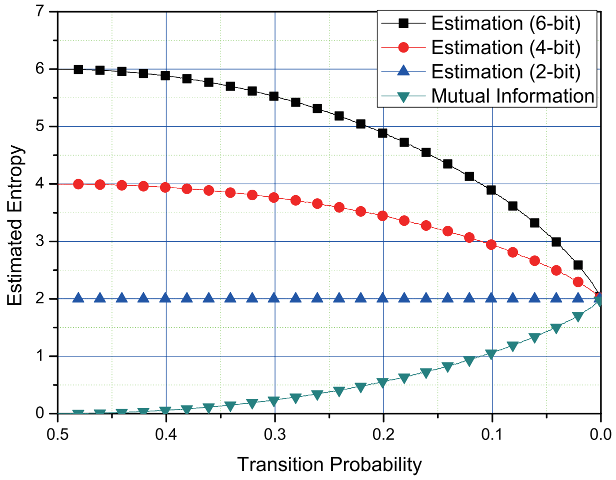

FIGURE 11. Estimated Mutual Information using Proposition 3.2 for $k = 2$ and 150,000 blocks. The estimated entropies for 6-bit/4-bit/2-bit block sizes are depicted at the same time.

entropy values $H(B_{i+2}B_{i+1}B_i)$ and $H(B_{i+1}B_i)$ are greater than 2 and $H(B_i)$ is equal to 1 (its maximum value) even for very small $\epsilon$. Therefore, the mutual information is close to 1 at $\epsilon \to 0$.

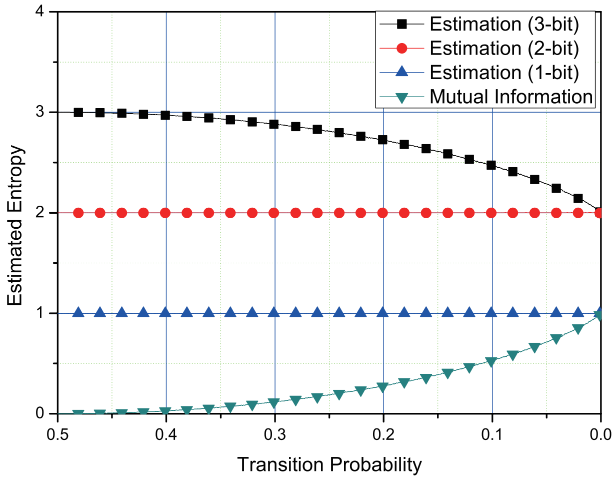

Figure 12. Estimated Mutual Information using Proposition 3.3 for $k = 1$ and 150,000 blocks. The estimated entropies for 3-bit/2-bit/1-bit block sizes are depicted at the same time.

## 5. Conclusions

In this paper, a new online test scheme for TRNGs is proposed based on the mutual information between consecutive $k$-bit random blocks since the mutual information is closely related to the amount of dependency between consecutive random blocks. It is shown that the mutual information can be estimated using two entropy values for the distinct sizes. Therefore, the number of occurrences of $2k$-bit blocks are counted as well as that of $k$-bit blocks for the estimation of the mutual information for this test. By using small amount of additional memory, the mutual information can be estimated as a measure oriented to the dependency of random data from TRNGs.

**Acknowledgement.** This work was supported by a grant from National Security Research Institute (NSRI), Korea (No. 2011-061) and the research fund from Chosun University, 2010.

## References

[1] M. Bucci and R. Luzzi, *Design of Testable Random Bit Generators*, in Proc. CHES 2005, LNCS, vol. 3659, pp. 147–156, 2005.

[2] J.-S. Coron, *On the security of random sources*, in Proc. PKC'99, LNCS, vol. 1560, pp. 29–42, 1999.

[3] ______, *An accurate evaluation of Maurer's universal test*, Selected areas in cryptography (Kingston, ON, 1998), 57–71, Lecture Notes in Comput. Sci., 1556, Springer, Berlin, 1999.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.

[5] W. Killmann and W. Schindler, *A proposal for: functionality classes and evaluation methodology for true (physical) random number generators*, AIS.31 Version 3.1, Sep. 25, 2001.

[6] U. M. Maurer, *A universal statistical test for random bit generators*, J. Cryptology **5** (1992), no. 2, 89–105.

[7] W. Schindler, *Efficient online tests for true random number generators*, in IACR Workshop on Cryptographic Hardware and Embedded Systems, CHES 2001, LNCS., vol. 2162, pp. 103–117, Springer, Berlin, 2001.

[8] B. Sunar, W. Martin, and D. Stinson, *A provably secure true random number generator with built-in tolerance to active attacks*, IEEE Trans. Computers **56** (2007), no. 1, 109–119.

[9] E. Trichina, et al., *Supplemental cryptographic hardware for smart cards*, IEEE Micro. **21** (2001), no. 6, 26–35.

[10] I. Vasyltsov, et al., *Fast digital TRNG based on metastable ring oscillator*, in Proc. CHES 2008, LNCS, vol. 5154, pp. 164–180, 2008.

Young-Sik Kim
Department of Information and Communication Engineering
Chosun University
Gwang 501-759, Korea
*E-mail address*: `iamyskim@chosun.ac.kr`

Yongjin Yeom
Department of Mathematics
Kookmin University
Seoul 136-702, Korea
*E-mail address*: `salt@kookmin.ac.kr`

Hee Bong Choi
The Attached Institute of ETRI
Daejeon 305-600, Korea
*E-mail address*: `gold@ensec.re.kr`