Fifteenth International Workshop on Algebraic and Combinatorial Coding Theory
June 18-24, 2016, Albena, Bulgaria *pp. 290–295*

On some properties of PRNGs based on block ciphers in counter mode

Alexey Urivskiy, Andrey Rybkin, Mikhail Borodin
urivskiy@infotecs.ru, alexey.urivskiy@mail.ru
andrey.rybkin@infotecs.ru, mikhail.borodin@infotecs.ru
JSC InfoTeCS, Moscow, Russia

Abstract. In this paper pseudorandom number generators based on block ciphers in counter mode are investigated. An idealized abstraction is applied which models the block cipher as a random permutation. We computed the number of output sequences of the idealized generators and estimated the conditional probability for the next element to appear given the prefix. In particular, we computed lower and upper bounds for that conditional probability.

1 Introduction

Deterministic pseudorandom number generators (PRNG) are important and useful cryptographic tools. Their construction can be dedicated or based on different cryptographic primitives [1], e.g. hash-functions or block ciphers. The PRNG we investigate is based on block ciphers in counter mode of operation [2], and it is considered as a candidate for standardization in Russia [3]. As a counterpart of it we also consider a well-known `CTR_DRBG` generator described in ISO/IEC 18031 [1] standard, which is based on one block cipher.

Since typically it is quite difficult to determine or even estimate the properties of real PRNGs, we focused on the properties of the constructions itself modeling block ciphers with random keys as random permutations. Under this assumption we were able to compute the number of different output sequences of the PRNGs, and estimate conditional probability for the next element of the generator to appear provided the previous elements of the sequence are known.

2 PRNGs Description

Let V_n be the set of all binary strings (vectors) of length n with the bitwise eXlusive OR addition defined on it. To every string $z_{n-1}||z_{n-2}||\ldots||z_0$ from V_n we put into one-to-one correspondence the integer $2^{n-1}z_{n-1}+\ldots+2z_1+z_0$. In what follows we do not distinguish strings and numbers.

Let $E(K,M)$ be a mapping of a block cipher encrypting a message $M \in V_n$ with a secret key $K \in V_k$, and let $E^{-1}(K,M)$ be the inverse decrypting mapping.

Consider two PRNGs, say **G1** and **G2**, based on block ciphers in counter mode. For both generators the counter *count* is initialized by a randomly and uniformly chosen $IV \in V_n$, and as the input an integer m and a secret key K are given. The output will be n-bit symbols $x_1, x_2, \ldots, x_m$.

G1: for i from 1 to m do:
$x_i := E(K, count)$; $count := (count + 1) \mod 2^n$.
This is essentially the widely used in practice `CTR_DRBG` generator [1].

G2 [3]: for i from 1 to m do:
$x_i := E(K, count) \oplus E^{-1}(K, (count + 1) \mod 2^n)$;
$count := (count + 1) \mod 2^n$.

The output sequences of both generators are evidently have a period of 2^n. In the following, we consider output sequences of length equal to this period.

3 PRNGs Properties

3.1 An Idealized Model for Generators

Let us idealize the generators by making two important assumptions.

Assumption 1. *Encryption (decryption) procedure of an n-bit block cipher with random key is modeled as a random permutation σ on V_n..*

Assumption 1 could hardly be strictly justified. The cardinality of the set of all permutations on V_n is $(2^n)!$, while a block cipher with k-bit key can give at most 2^k of them. However, this particular assumption is frequently used for proving the security of cryptographic constructions involving block ciphers.

Assumption 2. *Encryption and decryption procedures of a block cipher with the same random key are assumed to be independent, so they are considered to be two random and independent permutations σ_1 σ_2 on V_n.*

Block cipher encryption and decryption with the same key are the inverse operations, so they are dependent. However, the independence quite easily could be achieved, e.g. by taking two different ciphers or two independent keys.

In the rest of the paper, we call **G1** with applied Assumption 1 to be the generator **G1I**, while **G2** with both assumptions applied to be **G2I**.

Initializing Value. For **G2I** the permutations are chosen from the set of all permutations on V_n. Therefore **G2I** with permutations σ_1 and σ_2 and some initializing value IV could equivalently be considered as **G2I** with permutations σ_1^* and σ_2^*, and $IV = 0$:

$$\sigma_1^*(i) = \sigma_1\big((IV + i) \mod 2^n\big), \quad \sigma_2^*(i) = \sigma_2\big((IV + i + 1) \mod 2^n\big), \quad \forall i \in V_n.$$

The same is true for **G1I**. Thus the real IV for the idealized generators is not important, need not be kept secret, and can always be put to 0. Note that for

G1 and **G2** the situation is more complicated, since σ^* would not necessarily belong to the set of permutations defined by the block cipher.

3.2 G1I Properties

One of the most important properties of PRNGs is unpredictability of output symbols. For an ideal RNG with output alphabet of size N the conditional probability for the next symbol to appear is $\frac{1}{N}$ and coincides with the marginal probability since that generator is memoryless with all symbols are equally likely.

Output Sequences. Denote $|V_n| = 2^n = N$. Since an output sequence of **G1I** of length N consists of all values of any permutation on V_n, then the total number of different output sequences of **G1I** is exactly $N!$

Conditional Probability. Let us estimate conditional probability $P(a_{s+1}|a_s, a_{s-1}, \ldots, a_1)$ for a symbol a_{s+1} to appear provided s previously output symbols are known. Evidently, for $s = 1$ the probability is $\frac{1}{N}$.

For $s < N$ due to the bijectivity of permutations we have

$$P(a_{s+1}|a_s, a_{s-1}, \ldots, a_1) = \begin{cases} 0, & \text{if } a_{s+1} \in \{a_s, a_{s-1}, \ldots, a_1\}; \\ \frac{1}{N-s}, & \text{otherwise.} \end{cases}$$

3.3 G2I Properties

Output Sequences. Once again assume $|V_n| = 2^n = N$. For any permutation σ on V_n it holds that $\bigoplus_{i \in V_n} \sigma(i) = 0$. Then for any two permutations σ_1 and σ_2 it is true that

$$\bigoplus_{i \in V_n} (\sigma_1(i) \oplus \sigma_2(i)) = 0. \tag{1}$$

From (1) it follows that if we know any $N-1$ output symbols of **G2I**, then the N-th symbol can be computed as their bitwise sum. So **G2I** can output no more than N^{N-1} different sequences. To compute the exact number of sequences we use a theorem from [4], which for our notation can be reformulated as follows.

Theorem 1. *For any sequence of elements $b_0, b_1, \ldots, b_{N-1}$, $b_i \in V_n$, $i = \overline{0, N-1}$, $N = 2^n$, satisfying the condition*

$$\bigoplus_{i=0}^{N-1} b_i = 0, \tag{2}$$

there exists at least one pair of permutations σ_1, σ_2 on V_n such that $b_i = \sigma_1(i) \oplus \sigma_2(i)$.

From this theorem it immediately follows that **G2I** can output any sequence $b_0, b_1, \ldots, b_{N-1}$ satisfying (2). This in turn means that any $N-1$ out of N output symbols can take arbitrary values from V_n. So the total number of different output sequences of **G2I** is exactly N^{N-1}.

Equivalent Representation. To estimate conditional probability for **G2I** we give an equivalent description of how the next output symbol is generated.

Definition 1. *For any two permutations σ_1 and σ_2 on V_n the sequence $u_0, u_1, \ldots, u_{N-1}$, $u_i \in V_n$, is called the sum of σ_1 and σ_2 if*

$$u_i = \sigma_1(i) \oplus \sigma_2(i), \ \forall i = \overline{0, N-1}$$

Consider now a matrix $\mathbf{M}(i,j) = i \oplus j$, $i = \overline{0, N-1}$, $j = \overline{0, N-1}$.

Definition 2. *A sequence of pairs $(i_0, j_0), (i_1, j_1), \ldots, (i_{N-1}, j_{N-1})$, $i_l, j_l = \overline{0, N-1}$, $i_k \neq i_t$, $j_k \neq j_t$ for any $k \neq t$, is called a trajectory on matrix $\mathbf{M}$.*

Definition 3. *The sequence $\mathbf{M}(i_0, j_0), \mathbf{M}(i_1, j_1), \ldots, \mathbf{M}(i_{N-1}, j_{N-1})$ is called the output of the trajectory $(i_0, j_0), (i_1, j_1), \ldots, (i_{N-1}, j_{N-1})$.*

Proposition 1. *Between the set of ordered pairs of permutations on V_n and the set of trajectories on matrix $\mathbf{M}$ a one-to-one correspondence can be defined so that the sum of the pair of permutations will coincide with the output of the corresponding trajectory.*

Proof. Any two permutations σ_1 σ_2 on V_n induce the trajectory

$$(\sigma_1(0), \sigma_2(0)), (\sigma_1(1), \sigma_2(1)), \ldots, (\sigma_1(N-1), \sigma_2(N-1)).$$

Conversely, for any trajectory $(i_0, j_0), (i_1, j_1), \ldots, (i_{N-1}, j_{N-1})$ there exists a unique pair of permutations σ_1 and σ_2 such that

$$(i_0, j_0), \ldots, (i_{N-1}, j_{1-1}) = (\sigma_1(0), \sigma_2(0)), \ldots, (\sigma_1(N-1), \sigma_2(N-1))$$

Suppose that a sequence $u_0, u_1, \ldots, u_{N-1}$ is the sum of two permutations σ_1 and σ_2. Consider the output of the corresponding trajectory on $\mathbf{M}$:

$$\mathbf{M}(\sigma_1(0), \sigma_2(0)), M(\sigma_1(1), \sigma_2(1)), \ldots, \mathbf{M}(\sigma_1(N-1), \sigma_2(N-1)) =$$
$$= \sigma_1(0) \oplus \sigma_2(0), \sigma_1(1) \oplus \sigma_2(1), \ldots, \sigma_1(N-1) \oplus \sigma_2(N-1) = u_0, u_1, \ldots, u_{N-1}.$$

□

Proposition 1 gives us an equivalent description of how a symbol is generated by **G2I**. The next output symbol of **G2I** is some element selected from $\mathbf{M}$, moreover after the selection we strike out the row and column containing that element since only one element from any row and any column can be selected.

Note that the number of different trajectories on $\mathbf{M}$ is $(N!)^2$ while there are N^{N-1} different outputs of trajectories. For $N \geq 2$ it holds $(N!)^2 > N^{N-1}$, so there definetly exist different trajectories giving the same output.

Conditional Probability. Let us estimate $P(a_{s+1}|a_s, a_{s-1}, \ldots, a_1)$ for **G2I**. Consider the case $s < \frac{N}{2}$. The prefix $a_s, a_{s-1}, \ldots, a_1$ is an output of some partial trajectory $T(s) = (i_0, j_0), (i_1, j_1), \ldots, (i_{s-1}, j_{s-1})$. This means that some s rows and s columns were struck out from $\mathbf{M}$. Hence the next symbol a_{s+1} could be chosen from the remaining $(N-s) \times (N-s)$ matrix $\mathbf{L}$.

Any row (column) of $\mathbf{M}$ contains each symbol exactly once. So in the s struck out rows (columns) every symbol were deleted s times. Consider an $s \times s$ matrix $\mathbf{L}^*$ which lies on the intersection of the struck out columns and rows. Assume that the symbol a_{s+1} occurs ξ times in $\mathbf{L}^*$. Obviously $0 \le \xi \le s$. Therefore the number of a_{s+1} in $\mathbf{L}$ is $N - 2s + \xi$, and $P(a_{s+1}|T(s)) = \frac{N-2s+\xi}{(N-s)^2}$.

A fixed prefix can be obtained by different partial trajectories $T_t(s) = T_t$ on $\mathbf{M}$ each of which leads to its own $\mathbf{L}$. Since all the events of having T_t are disjoint, we obtain $P(a_1, \ldots, a_s) = \sum_t P(T_t)$, where the sum is over all T_t giving $a_1, \ldots, a_s$ as an output. From Bayes' rule we get:

$$P(a_{s+1}|a_s, \ldots, a_1) = \frac{P(a_{s+1}, \ldots, a_1)}{P(a_s, \ldots, a_1)} = \frac{\sum\limits_t P(a_{s+1}, T_t)}{\sum\limits_t P(T_t)} = \frac{\sum\limits_t P(a_{s+1}|T_t)P(T_t)}{\sum\limits_t P(T_t)}.$$

It is unknown how many different T_t lead to $a_s, \ldots, a_1$. However, from the above formula it follows that as upper and lower bounds on $P(a_{s+1}|a_s, \ldots, a_1)$ we can take the upper and lower bounds on $P(a_{s+1}|T(s))$ bounding ξ. Thus,

$$P_1 = \frac{N-2s}{(N-s)^2} \le P(a_{s+1}|a_s, a_{s-1}, \ldots, a_1) \le \frac{N-s}{(N-s)^2} = P_2.$$

Let us estimate how many different a_{s+1} might have the conditional probability equal to P_1. Consider an $(N-s) \times s$ matrix $\mathbf{Z}$ which is the struck out columns except the matrix $\mathbf{L}^*$. The symbols corresponding to P_1, say there are z_1 of them, are those which occur in $\mathbf{Z}$ exactly s times. Trivially there are $z_1 \le \frac{(N-s)s}{s} = N - s$ such symbols.

Estimate now how many different a_{s+1} might have the conditional probability equal to P_2. The symbols corresponding to P_2, say there are z_2 of them, are those which do not occur in $\mathbf{Z}$ at all. Therefore $\mathbf{Z}$ is filled with at most $N - z_2$ other symbols each of which can occur no more than s times. So it must be $(N - z_2)s \ge (N - s)s$, and $z_2 \le s$.

4 Generators Comparison and Conclusion

4.1 Comparison

Sequences. The **G1I** generator can output $N!$ different sequences, **G2I** can output N^{N-1} sequences, while an ideal RNG could output all N^N possible

sequences. Using Stirling's approximation we obtain

$$\frac{N!}{N^{N-1}} \approx \frac{\sqrt{2\pi N} \cdot N^N}{e^N} \frac{N}{N^N} = \frac{N\sqrt{2\pi N}}{e^N} = e^{-(N - \ln N\sqrt{2\pi N})}.$$

So concerning the output sequences **G2I** is far better than **G1I**, and is close to an ideal RNG. Moreover, up to the length $N-1$ **G2I** can output any sequence.

Probabilities. The deviation of conditional probability for the next output symbol to appear given a prefix from $1/N$, which corresponds to an ideal RNG, is an important cryptographic property of PRNGs. For **G1I** after s steps this probability takes two values 0 and $\frac{1}{N-s}$, while the latter value corresponds to exactly $N-s$ symbols.

The **G2I** generator is closer to an ideal one: for $s < \frac{N}{2}$ the conditional probability can take values from $P_1 = \frac{N-2s}{(N-s)^2}$ (for no more than $N-s$ symbols) up to $P_2 = \frac{1}{N-s}$ (for no more than s symbols). Observe also that $P_2 - \frac{1}{N} = \frac{s}{N(N-s)}$ and $\frac{1}{N} - P_1 = \frac{s^2}{N(N-s)^2}$. Therefore $P_1 < \frac{1}{N} < P_2$ for any s.

4.2 Conclusion

In this paper, we considered PRNGs based on block ciphers working in counter mode of operation. To evaluate the properties of the PRNGs we idealized them by replacing block ciphers with a random secret key by random permutations. And at the same time we considered the encryption and decryption procedures to be independent. The qualitative conclusion is that the idealized PRNG on two block ciphers reveals better characteristics in terms of the number of output sequences and conditional probability for the next symbol to appear than those on one block cipher. This clearly justifies the doubled computational complexity required to produce the next symbol in the two-cipher generator.

References

[1] International standard ISO/IEC 18031:2011, Information technology — Security techniques — Random bit generation.

[2] International standard ISO/IEC 10116:2006, Information technology — Security techniques — Modes of operation for an n-bit block cipher.

[3] Lavrikov I., Rudskoy V., On approaches to design of key derivation mechanisms and pseudorandom sequence generators. — *Proceedings of Ruscrypto'2016.* — 2016 (in Russian).

[4] Hall M., A combinatorial problem on abelian groups, *Proceedings of the American Mathematical Society.* — 1952. — Vol. 3., N. 4. — pp. 584–587.