# The Randomness Property of a Random Sequence Based on Occurrence Probability of an *m*-bits Pattern after Insertion Attack with NOL and OL Pattern Approaches

[1]Santi Indarjani,[2]Kiki A. Sugeng, and [3]Belawati H. Widjaja

***Abstract*— Since random numbers are required in a large number of cryptographic applications, e.g. as session key, IV, ephemeral key, nonce, or challenges in zero knowledge protocols, etc., then its failure to fulfill the randomness property could be a potential weakness for the system which employed those cryptographic applications to deliver its security services. The previous researches on insertion attack toward random binary sequences produced by five PRNG algorithms have indicated that some postattack random sequences failed to pass the statistical randomness tests. As an addition, some postattack random sequences are potentially being distinguished from the preattack (target) sequence under advantage value $\varepsilon = 0.00001$ and $\varepsilon = 0.0001$ for statistical distance test and entropy different test respectively. This meant that the insertion attack is potential to change the distribution of each pattern occurs in the post attack binary sequence, that could cause the lost of randomness. The effects is also vary for each algorithm. Based on those findings, then we extend the research to measure the effects of insertion attack on a random sequence theoretically by examining the occurrence probability of each *m*-bits pattern after the attack with non-overlapping (NOL) patterns and overlapping (OL) pattern approaches. The results showed that the insertion attack will cause the bias from uniformity in the postattack sequence when the insertion attack is not balance, the other hand the distribution each *m*-bits pattern remains uniform that potentially keeps the randomness property of the target sequence.**

***Index Terms*—randomness, insertion attack, *m*-bit pattern, occurrence probability, NOL pattern, OL pattern**.

## I. INTRODUCTION

In cryptographic application, randomness is a very important element that is used to ensure the security of the system. As Kerckhoff Principles said that the strength of a cryptographic algorithm entirely depends on the secrecy of the key, where the adversary is assumed has complete knowledge about the crypto algorithm except the secret key [1]. By using random variable for certain critical parameters such as key, seed, nonce, IV or others, it will strengthened the system where adversary will be forced to maximum efforts in finding the secret values.

Some related researches in ([2], [3], [4]) mentioned that some lacks in generating random values will cause the security system failed. On 11 August 2013, BitCoin Foundation even announced that a certain component in Android that responsible to produce random sequences contained some bugs, which caused all Wallet Applications such as BitCoin Wallet is vulnerable against thieves. [5]. The fact confirmed that it is important that the random numbers produced by an RNG/a PRNG is ensured to be random.

The empirical studies of insertion attack effects on Mother of all random number generators [6] and AES-based pseudorandom number generator [7], have proved that the attack caused some samples failed to pass the randomness tests after the attack under significant level of $\alpha = 0.01$. Meanwhile [8] found the effects on four PRNGs ANSI X9.17, ANSI X9.31, Dragon, and Rabbit algorithm caused some postattack sequences failed the NIST randomness tests under level significant of $\alpha = 0.001$. Table I showed the different effects on the five algorithms.

TABLE I
NIST RANDOMNESS TEST RESULTS ON FIVE ALGORITHMS ON $\alpha = 0.01$

| | AES-128 | | | AES-192 | | | AES-256 | | | ANSI | | DRAGON | RABBIT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OFB | CFB | CTR | OFB | CFB | CTR | OFB | CFB | CTR | X9.17 | X9.31 | | |
| Total failed experiment | 3 | 2 | 3 | 3 | 1 | 7 | 5 | 2 | 3 | 4 | 5 | 2 | 3 |
| Total failed tests | 4 | 3 | 3 | 3 | 1 | 8 | 8 | 2 | 3 | 6 | 6 | 2 | 5 |

Using statistical distance test, it is also indicated that the postattack random sequences from AES-based PRNG still could not be distinguished from the preattack random sequences under advantage $\varepsilon = 0.01$ [7]. The effects would be significant under $\varepsilon = 0.00001$ from further experiments. Moreover, from [9] and [10] it is found that the indistinguishability of four PRNGs are significant under advantage value $\varepsilon = 0.00001$ based on statistical distance tests and $\varepsilon = 0.0001$ based on entropy different tests, as is shown in Table II.

[1] Santi Indarjani is with the National Crypto Institute, Bogor, Indonesia (email : santi.indaryani@ui.ac.id, santi.indarjani@stsn-nci.ac.id).

[2]Kiki A. Sugeng is with Mathematic Department, Faculty of Mathematic and Natural Sciences, University of Indonesia, Depok, Indonesia (email : kiki@sci.ui.ac.id).

[3]Belawati H. Widjaja is with Faculty of Computer Science, University of Indonesia, Depok, Indonesia (email :bela@cs.ui.ac.id).

TABLE II
TEST RESULTS OF DISTINGUISHED SEQUENCES FROM ORIGINAL (%)

| Advantage value | Bit Level | Max statistical distance test | | | | Max entropy difference test | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Rabbit | Dragon | X9.17 | X9.31 | Rabbit | Dragon | X9.17 | X9.31 |
| ε = 0,0001 | 1-bit | 0 | 0 | 0 | 0 | 40 | 36 | 36 | 0 |
| | 2-bit | 0 | 0 | 0 | 0 | 28 | 20 | 24 | 0 |
| | 3-bit | 0 | 0 | 0 | 0 | 32 | 40 | 24 | 0 |
| ε = 0,00001 | 1-bit | 80 | 80 | 76 | 76 | 80 | 68 | 72 | 44 |
| | 2-bit | 64 | 60 | 60 | 64 | 64 | 56 | 60 | 32 |
| | 3-bit | 68 | 64 | 68 | 64 | 68 | 72 | 64 | 32 |

Based on those results, the effects on five PRNG ANSI X9.17, X9.31, Rabbit, Dragon, and AES-based PRNG are quite similar against the insertion attack with random bits under statistical distance test meanwhile ANSI X9.31 is stronger against the attack under entropy different test. This fact drives a curiosity to find out more about the effects of insertion attack on a random/pseudorandom sequence.

In the randomness test tools such as Five Basic Tests [11], NIST Randomness Tests [12] or Die Hard Tests [13], there are some tests that are measuring the randomness property by examining the occurrence of every bit-pattern in the sequences. E.g., Frequency test is used to measure the occurrence of bit 1 and 0 in the sequence. The sequence with length *n*-bits will pass the frequency test if the total number of bit 0 and bit 1 is equal when *n* is even, or different at most 1 bit when *n* is odd. In other words, the expected probability of bit 0 and bit 1 to occur in the sequence is equal such that $P(0) = P(1) \approx \frac{1}{2}$. Then Serial test is used to meassure the occurrence of 2-bits pattern 00, 01, 10 and 11 in the sequence. Each 2-bits pattern is expected equally likely to occur in the sequence with probability close to $\frac{1}{4}$. Each 2-bits pattern is expected to occur in the sequence with probability close to 1/4. In Runs test, the occurrence of n consecutive bits (which is called as "gap" for bit 0s or "block" for bit 1s) is expected equally likely in the sequence. Globally, on a random sequence, it is expected that every possible *m*-bits pattern, where *m* = 1,2,…, *n*, occurred in the sequence with the same (or close) probability.

Under considerations above, we proceed to explore the insertion attack effects on a random/pseudorandom sequence by measuring the probability of each pattern occurred in the postattack sequence. The preattack binary sequence is assumed to be random or already pass the statistical test for randomness, where the distribution of the sequence is uniform based on [12]. In other words, every pattern in the sequence has the same (close) probability to occur or we define it as balance. To see all the possible effects, we examine all possible ways of insertion attack with non-overlapping (NOL) pattern and overlapping (OL) pattern approaches. The observations are conducted on the insertion attack with random bits and extreme bits.

The results showed that the effects of the insertion attack is depending on the way of the insertion take place, the position of the insertion, and also the balance of the inserted bits. From theoretical proofs we observed, globally there are four principle results:

1) The insertion attack of balance *m*-bits pattern will definitely produce the balance of the *m*-bits pattern in the postattack sequence which potentially maintains the randomness property of the sequence.
2) Contrary, the insertion attack of the unbalance *m*-bits pattern will cause the occurrence of the bits patterns in the postattack sequence is unbalance, indicating by the bias of the occurrence probability of an *m*-bits pattern from $\frac{1}{2^m}$ which leads the sequence to lose the randomness property.
3) The bias of the attack effects with NOL pattern approaches on the position *nm*, *n* = 1,2,... can be generalized. Meanwhile the bias attack effects with NOL approach on position *nm*+*t*, *n* = 1,2,... and *t* = 1,2, …, (*m*-1) or with OL approach cannot be generalized because the attack will produce the new different patterns for each case that depends on the bits after and before the insertion point.
4) The insertion attack with extreme bits causes a significant bias of the attack effects that trivially defects the randomness property of the sequence.

To make it clear, we divide this paper in 5 sections includes: Introduction, Background Theory, Insertion Attack Effects with Random Bits, Insertion attack effects with extreme bits, and Conclusion. Section 3 is divided into 4 subsections i.e.: Notation, 1-bit insertion attack effect, *m*-bit insertion attack with NOL pattern approach, and *m*-bit insertion attack with OL pattern approach.

## II. BACKGROUND THEORY

### *A. Randomness and unpredictability*

A series of random bits could be interpreted as the result of flipping an unbiased "fair" coin with two sides that are labeled "0" and "1", in which each side has probability ½, with the provisions of each coin flip is independent and mutually disjoint. If these conditions are fulfilled then the coin flip n times will result unbiased random bit stream, where the value of "0" and "1" will be distributed randomly or uniformly distributed. [12].

There are two types of random numbers. First is a truly random number that is produced by a random number generator called as an RNG. RNG uses a non-deterministic source (i.e., the entropy source), along with some processing function (i.e., the entropy distillation process) to produce randomness. Second is a pseudorandom number which is produced by a pseudorandom number generator called as a PRNG. A PRNG uses one or more inputs (called "seed") and generates multiple "pseudorandom" numbers. The outputs of a PRNG are typically deterministic functions of the seed which are based on mathematical formulations. This gives the output of the PRNG will be reproduced by using the same seed, contrary with the RNG that could not be reproduced. [14].

Random and pseudo random numbers generated for cryptographic application should be unpredictable. On a PRNG, if the seed is unknown then the following numbers in the output sequence should be unpredictable for any knowledge of the previous bits. This property is called forward unpredictability. It is also required that based on the

knowledge of some values of the output bits, it should be difficult to determine the value of the seed, which is called as the backward nature of unpredictability. In this case there should be guaranteed no correlation between seed and output bits generated by the seed, where each element of the output bits should have probability ½. [12]

From [12], it is stated that the sequences to be tested is considered to meet the assumptions:

- Uniformity: At any point in the generation of a sequence of random or pseudorandom bits, the occurrence of a zero or one is equally likely, i.e., the probability of each is exactly 1/2. The expected number of zeros (or ones) is $n/2$, where $n$ = the sequence length.
- Scalability: Any test applicable to a sequence can also be applied to subsequences extracted at random. If a sequence is random, then any such extracted subsequence should also be random. Hence, any extracted subsequence should pass any test for randomness.
- Consistency: The behavior of a generator must be consistent across starting values (seeds).So that it is inadequate to test a PRNG based on the output from a single seed, or an RNG on the basis of an output produced from a single physical output.

*B. Probability Theory*

Definition 1[15]

Two events $E$ and $F$ are disjoint if there are no outcomes common to both E and F which is notated as $E \cap F = \emptyset$.

$E \cup F$ is the collections of all outcomes in either $E$ or F so that the probability of $E \cup F$ is the sum of each probability $E$ and $F$ that is written as

$$P(E \cup F) = P(E0 + P(F) \quad (1)$$

If $E$ and $F$ are not disjoint, then the probability of the event $E \cup F$ is not the sum of P($E$) and P($F$), because the outcomes common to both $E$ and $F$ should not be counted twice which is formulated as

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) \quad (2)$$

Definition 2 [15]

Two events $E$ and $F$ are said to be independent if the probability of both occurrence are the product of their individual probability, notated as

$$P(E \cap F) = P(E)P(F) \quad (3)$$

*C. M-bits pattern*

Bit patterns can be defined as all possible forms of $n$ bits. E.g., one bit represents two patterns, i.e., 1 and 0, 2-bits represents 4 patterns, i.e., 00, 01, 10, and 11, then 3 bits represents 8 patterns, i.e., 000,001, 010, 011, 100, 101, 110, and 111, and so on. Globally, $n$-bits will represent $2^n$ patterns. As mentioned above, a sequence is considered as random if all of possible patterns are equally likely to occur in the sequence. Generally, each $m$-bits pattern is expected has the same probability $1/2^m$ of occurrence.

To examine the $m$-bits pattern that occurred in a random sequence with length $n$-bits and $m > 1$, we use two different approaches, i.e., non-overlapping pattern (NOL) and overlapping pattern (OL). In NOL approach, the patterns are considered from the sequence without overlapping. Meanwhile, in OL approach, the patterns are considered from the sequence with overlapping.

For example, from a random sequence of 24-bits 011010001101001110, there are 6 forms of 3-bits pattern considered with NOL approach i.e. 011, 010, 001, 101, 001, and 110. The other hand, there are 24 patterns for cyclic overlapping approach i.e. 011, 110, 101, 010,…,110, or 22 patterns for non-cyclic overlapping approach by omitting the last pattern.

Generally, there are $n/m$ forms of an $m$-bits pattern can be derived from an $n$-bits sequence with NOL pattern. Meanwhile with OL pattern there is $n$ forms for cyclic OL and $n$-($m$-1) forms for non-cyclic OL.

In this research, we use NOL and cyclic-OL approach in examining the patterns, under consideration that a binary sequence produced by a PRNG is periodic. The occurrence probability of an $m$-bit pattern with NOL approach and cyclic-OL approach are proved to be the same as presented below.

Let $n_{i_1 i_2 \ldots i_m}$ represents the $m$-bits pattern that occurs in a sequence with length $n$ where $i_1 i_2 \ldots i_m \in \{0,1\}$ so that the occurrence probability of $n_{i_1 i_2 \ldots i_m}$ in the sequence notated as:

$$P\left(n_{i_1 i_2 \ldots i_3}\right) = \frac{n_{i_1 i_2 \ldots i_3}}{N} \quad (4)$$

$N$ is the total number of $m$-bits pattern in the sequence, for which, $N = n/m$ for NOL pattern, $N = n$ for cyclic-OL pattern, and $N = n - (m\text{-}1)$ for non-cyclic OL pattern.

From probability principle we know the sum of all individual probability of each pattern is 1. Suppose that each pattern has the same probability $1/2^m$ to be occurred. So the occurrence probability of each $m$-bits pattern in the sequence of length $n$ with NOL pattern can be determined as in the Eq. 5.

$$P\left(n_{i_1 i_2 \ldots i_m}\right) = \frac{P(n'_{i_1 i_2 \ldots i_m})N}{N} \approx \frac{\frac{1}{2^m}\left(\frac{n}{m}\right)}{\frac{n}{m}} = \frac{1}{2^m} \quad (5)$$

The other hand, the occurrence probability of each $m$-bits pattern in the sequence with cyclic OL pattern can be determined as in the Eq. 6.

$$P\left(n_{i_1 i_2 \ldots i_m}\right) = \frac{P(n'_{i_1 i_2 \ldots i_m})N}{N} \approx \frac{\frac{1}{2^m}(n)}{n} = \frac{1}{2^m} \quad (6)$$

From Eq. 5 and 6 above, we see that the occurrence probability of each $m$-bits pattern in a random sequence with length $n$ has the same probability to occur i.e. $1/2^m$ for NOL pattern or cyclic OL pattern approach.

On the next sections we will show our proofs about the effects of insertion attack based on the occurrence probability of each $m$-bits pattern in a random sequence with length $n$-bits after the attack. The original sequence produced by an RNG/ a PRNG as a target sequence is defined as a preattack sequence where is assumed as

random. And the target sequence after insertion attack is defined as a postattack sequence.

## III. Insertion Attack Effect with Random Bits

In this section we present the effect of insertion attack with random bits against a binary random sequence by measuring the occurrence probability of each *m*-bit pattern after the attack. It is assumed that the preattack sequence is random, which means that each pattern in the sequence has the same (close) probability to occur, which is defined as balance in this paper.

There are two possible conditions of the bits are inserted into the sequence produced by an RNG/ a PRNG. First the bits inserted are balance, and second the bits inserted are not balance. The measurements of the attack effect are conducted under the two conditions.

### A. *Notation*

Some notations used in this paper are:

$U = u_1u_2 \dots u_n$ is the preattack sequence from RNG/PRNG

$V = v_1v_2 \dots v_n$ is the sequence of bits inserted into *U*.

$n_{i_1i_2\dots i_m}$ is the *m*-bit pattern in the preattack sequence *U*.

$s_{i_1i_2\dots i_m}$ is the *m*-bit pattern that is inserted into *U*.

$n'_{i_1i_2\dots i_m}$ is the *m*-bit pattern in postattack sequence.

$P(n_{i_1i_2\dots i_m})$ is the occurrence probability of $n_{i_1i_2\dots i_m}$ in the preattack sequence *U*.

$P(s_{i_1i_2\dots i_m})$ is the probability of $s_{i_1i_2\dots i_m}$ in *V* to be inserted into *U*.

$P(n'_{i_1i_2\dots i_m})$ is the occurrence probability of $n_{i_1i_2\dots i_m}$ in the postattack sequence.

### *B. 1-bit insertion attack effect*

The insertion attack effects is considerend under two conditions. First, where the bits inserted are balance (Case 1), and second, where the bits inserted are not balance (Case 2 and 3). The sequence of bits inserted has length *s* and the target target sequence *U* has length *n* bits, so that the postattack sequence *U*+*V* has length *n*+*s* bits.

Case 1:

Suppose the bits that are inserted in the sequence are balance that means the probability of bit 0 and bit 1 are the same, i.e. $P(s_0) = P(s_1) \approx \frac{1}{2}$. Since each event of 1-bit insertion attack are independent, so that the probability of bit 0 (or 1) to occur in the postattack sequence *U*+*V* can be defined as

$$\begin{aligned} P(n'_0) &= P(n_0) + P(s_0) \\ &\approx \frac{\frac{1}{2}n+\frac{1}{2}s}{n+s} = \frac{n+s}{2}.\frac{1}{n+s} = \frac{1}{2} \end{aligned} \quad (7)$$

Case 2:

Suppose that the probability of bit 0 to be inserted in a sequence is bigger than ½. Let $P(s_0) \approx \frac{1}{2} + \delta$, $0 < \delta < 1$, such that $P(s_1) < \frac{1}{2}$ where $P(s_0) + P(s_1) = 1$. Then probability of bit 0 to occur in the postattack sequence can be defined as

$$\begin{aligned} P(n'_0) &= P(n_0) + P(s_0) \approx \frac{\frac{1}{2}n+\left(\frac{1}{2}+\delta\right)s}{n+s} \\ &= \frac{\frac{1}{2}n+\frac{1}{2}s+\delta s}{n+s} = \frac{\frac{1}{2}(n+s)+\delta s}{n+s} \\ &= \frac{1}{2} + \frac{\delta s}{(n+s)} > \frac{1}{2} \end{aligned} \quad (8)$$

Case 3:

In contrary with case 2, the probability of bit 0 for being inserted is less than bit 0. Let $P(s_0) \approx \frac{1}{2} - \delta, 0 < \delta < 1$, such that $P(s_1) > \frac{1}{2}$ where $P(s_0) + P(s_1) = 1$. Then with the same way, we could define the occurrence probability of bit 0 in the postattack sequence as

$$\begin{aligned} \mathrm{P}(n'_0) &= P(n_0) + P(s_0) \\ &\approx \frac{\frac{1}{2}n+\left(\frac{1}{2}-\delta\right)s}{n+s} = \frac{\frac{1}{2}n+\frac{1}{2}s-\delta s}{n+s} \\ &= \frac{\frac{1}{2}(n+s)-\delta s}{n+s} = \frac{1}{2} - \frac{\delta s}{(n+s)} \end{aligned} \quad (9)$$

From the result above we propose the Theorem 1 about 1-bit insertion attack effects on a random sequence.

**Theorem 1 (1-bit insertion attack)**

Suppose $U = u_1u_2 \dots u_n$ is a random sequence from a RNG/PRNG with length *n* bits, and $V = v_1v_2 \dots v_s$ is a sequence of *s* bits to be inserted into *U* in any position. Then we found:

1. The occurrence probability of bit 0 (or bit 1) in the postattack sequence is $\approx \frac{1}{2}$ if only if bit 0 and bit 1 in *V* has the same probability to be inserted into *U*.
2. The probability of bit 0 (or bit 1) to occur in the postattack sequence will bias from 1/2 with error $\pm\frac{\delta s}{(n+s)}$ if only if the probability of bit 0 and probability of bit 1 to be inserted are not balance. The sign depends on the probability of bit 0 (or 1) to be inserted in *U*.

Proofs:

1. The proof of the right side is clear as shown in the equation (7). On the other hand, let assume the occurrence probability of bit 0 (or bit 1) after the attack is $\approx \frac{1}{2}$ and let the probability of bit 0 (or bit 1) to be inserted in the random sequence *U* is $\left(\frac{1}{2}+\delta\right) \not\approx \frac{1}{2}$. Based on mutually exclusive property, then it is trivial that the probability of bit 0 (or 1) in the preattack sequence *U* will be $\left(\frac{1}{2}-\frac{\delta}{2n}\right) \not\approx \frac{1}{2}$. It is contradictive with the basic assumption that before the attack, probability of bit 0 (or 1) in a random sequence *U* to be occur is $\approx \frac{1}{2}$. Therefore it is concluded that the probability of bit 0 (or 1) in *U* before attack will be $\approx \frac{1}{2}$ if only probability of bit 0 in *V* is $\approx \frac{1}{2}$. It proved that the first condition holds.
2. Let say that the probability of bit 0 (or bit 1) to be inserted is $\frac{1}{2}+\delta$ so that probability of bit 1 is less than $\frac{1}{2}$ (or vice versa). The proof of the second condition for the right side is clearly shown in (8) and (9). On the other hand, the proof for the left side is taken by Contradiction. Let the probability of bit 0 to occur after the attack is $\left(\frac{1}{2}+\delta\right) > \frac{1}{2}$ and let the probability of bit 0

to be inserted is $\approx \frac{1}{2}$ then based on mutually exclusive property we will have the probability of bit 0 to occur in the preattack sequence $U$ is $\left(\frac{1}{2} + \frac{\delta(n+s)}{2n}\right) \not\approx \frac{1}{2}$. It is contradictive with the basic assumption that the preattack sequence $U$ is random so that the occurrence probability of bit 0 and 1 is ≈ ½. Therefore it is concluded that the occurrence probability of bit 0 after the attack is $\not\approx \frac{1}{2}$ if only if the occurrence probability of bit 0 to be inserted is $\not\approx \frac{1}{2}$. It proved that the theorem point 2 holds.

**Corollary 1:**

The 1-bit insertion attack is potential to reduce or damage the randomness property of a random sequence if the probability of bit 1 and 0 to be inserted is not balance.

As mentioned above for pattern $m > 1$, there are two approaches of constructing the pattern, i.e. with NOL and OL approach. The measurements of the attack effects for pattern $m > 1$ will be divided into two categories as described in sub section C and D.

*C. m-bit insertion attack with NOL pattern approach*

For $m > 1$ with NOL approach, there are two possible conditions. First, the insertion attack are conducted in the preattack sequence at the position of $nm$, $n$ = 1,2,3,.... Second, is when the bits are inserted on position $nm+t$, $n$ = 1,2,3, ....; $t$ = 1,2, ..., ($m$-1). The illustration of the two conditions for pattern $m$ = 2 bits can be seen in Fig 1 and Fig. 2.

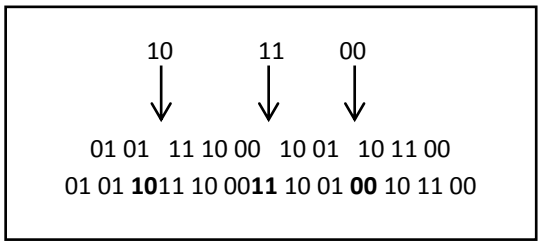


Fig 1.2-bits insertion attack on position *nm* with NOL approach

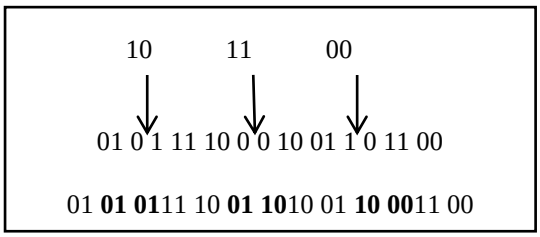


Fig.2. 2-bits insertion attack on position *n2+t* with NOL approach

*1) On position nm, n = 1,2,3, ...*

First we discuss about the insertion attack effects on $m$ = 2 for NOL approach on the position $nm$, $n$ = 1,2,3, ...

From Fig. 1 we can see that the insertion on position $nm$, $n$ = 1,2,3, ... does not change the patterns in target sequence before and after the insertion point, but only changes the total number of some $m$-bits patterns in the postattack sequence based on the distribution of the $m$-bits pattern in the inserted bits. The probability of each pattern to occur in the postattack sequence is measured in 3 possible conditions as discussed below.

Case 1:

Suppose each 2-bits pattern in sequence $V$ has the same probability to be inserted into sequence $U$. Let $P(s_{00}) = P(s_{01}) = P(s_{10}) = P(s_{11}) \approx 1/4$, then the probability of pattern bit 00 (or other pattern) to occur in the postattack sequence can be determined as

$$\begin{aligned} P(\mathrm{n}'_{00}) &= P(n_{00}) + P(s_{00}) \\ &\approx \frac{\frac{1}{4}\left(\frac{n}{2}\right)}{\frac{n+s}{2}} + \frac{\frac{1}{4}\left(\frac{s}{2}\right)}{\frac{n+s}{2}} \\ &= \frac{\frac{1}{4}\left(\frac{n+s}{2}\right)}{\frac{n+s}{2}} = \frac{1}{4} \end{aligned} \quad (10)$$

Case 2:

Suppose the probability of an 2-bits pattern is bigger than the expected probability ¼. Let $P(s_{00}) \approx \frac{1}{4} + \delta$ such that $P(s^c{}_{00}) < \frac{3}{4}$ where $P(s^c{}_{00}) = P(s_{01}) + P(s_{10}) + Ps_{11})$ and $P(s_{00}) + P(s^c{}_{00}) = 1$. Then the probability of pattern 00 to occur in the postattack sequence is

$$\begin{aligned} P(\mathrm{n}'_{00}) &= P(n_{00}) + P(s_{00}) \\ &\approx \frac{\frac{1n}{42} + \left(\frac{1}{4} + \delta\right)\frac{s}{2}}{\frac{n+s}{2}} \\ &= \frac{\frac{n}{8} + \frac{s}{8} + \frac{\delta s}{2}}{\frac{n+s}{2}} = \frac{n+s+4\beta}{4(n+s)} \\ &= \frac{1}{4} + \frac{\beta}{(n+s)} \end{aligned} \quad (11)$$

Case 3:

In contrary with case 2, let $P(s_{00}) \approx \frac{1}{4} - \delta < \frac{1}{4}$ so that $P(s^c{}_{00}) > \frac{3}{4}$ where $P(s^c{}_{00}) = P(s_{01}) + P(s_{10}) + Ps_{11})$ and $P(s_{00}) + P(s^c{}_{00}) = 1$. Then the probability of pattern bit 00 to occur in the postattack sequence is :

$$\begin{aligned} P(\mathrm{n}'_{00}) &= P(n_{00}) + P(s_{00}) \\ &\approx \frac{\frac{1n}{42} + \left(\frac{1}{4} - \delta\right)\frac{s}{2}}{\frac{n+s}{2}} \\ &= \frac{\frac{n}{8} + \frac{s}{8} - \frac{\delta s}{2}}{\frac{n+s}{2}} = \frac{n+s-4\beta}{4(n+s)} \\ &= \frac{1}{4} - \frac{\beta}{(n+s)} \end{aligned} \quad (12)$$

From the proofs of 2-bits pattern above, the attack on the preattack sequence on position even, does not change the distribution of 2-bits patterns in the postattack sequence when the bits inserted are balance. In this case the randomness property of the target sequence is possible to be kept.

On the other hand, when the inserted bits are not balance, the attack will affect the distribution of each 2-bits pattern to occur in the postattack sequence. In other words, the attack can change the distribution of 2-bits patterns in the postattack sequence that potential to reduce or damage the randomness property of the target sequence. Those results are the same with the case in 1-bit insertion attack.

With the same way it can be proved that the condition above can be generalized for $m$-bits pattern as is described below.

Case1:

For $P(s_{i_1 i_2 \dots i_m}) = 1/2^m$ and $i_1 i_2 \dots i_m \epsilon \{0,1\}$ then the probability of pattern $n'_{i_1 i_2 \dots i_m}$ in the postattack sequence can be defined as

$$\begin{aligned} P(n'_{i_1 i_2 \dots i_m}) &= P(n_{i_1 i_2 \dots i_m}) + P(s_{i_1 i_2 \dots i_m}) \\ &\approx \frac{\frac{1}{2^m}\frac{n}{m} + \frac{1}{2^m}\frac{s}{m}}{\frac{n+s}{m}} \\ &= \frac{\frac{1}{2^m}\left(\frac{n+s}{m}\right)}{\frac{n+s}{m}} = \frac{1}{2^m} \end{aligned} \quad (13)$$

Case 2:

For $P(s_{i_1 i_2 \dots i_m}) > 1/2^m$ where $P(s_{i_1 i_2 \dots i_m}) \approx \frac{1}{2^m} + \delta$ and $i_1 i_2 \dots i_m \epsilon \{0,1\}$, then the probability of pattern $n'_{i_1 i_2 \dots i_m}$ in the postattack sequence can be defined as

$$\begin{aligned} P(n'_{i_1 i_2 \dots i_m}) &= P(n_{i_1 i_2 \dots i_m}) + P(s_{i_1 i_2 \dots i_m}) \\ &\approx \frac{\frac{1}{2^m} \cdot \frac{n}{m} + \left(\frac{1}{2^m} + \delta\right)\frac{s}{m}}{\frac{n+s}{m}} \\ &= \frac{(n+s)+2^m(\delta s)}{2^m(n+s)} \\ &= \left[\frac{1}{2^m} + \frac{\delta s}{(n+s)}\right] \end{aligned} \quad (14)$$

Case 3:

For $P(s_{i_1 i_2 \dots i_m}) < 1/2^m$ where $P(s_{i_1 i_2 \dots i_m}) \approx \frac{1}{2^m} - \delta$ and $i_1 i_2 \dots i_m \epsilon \{0,1\}$, then the probability of pattern $n'_{i_1 i_2 \dots i_m}$ in the postattack sequence can be defined as

$$\begin{aligned} P(n''_{i_1 i_2 \dots i_m}) &= P(n_{i_1 i_2 \dots i_m}) + P(s_{i_1 i_2 \dots i_m}) \\ &\approx \frac{\frac{1}{2^m} \cdot \frac{n}{m} + \left(\frac{1}{2^m} - \delta\right)\frac{s}{m}}{\frac{n+s}{m}} \\ &= \frac{(n+s)-2^m(\delta s)}{2^m(n+s)} \\ &= \left[\frac{1}{2^m} - \frac{\delta s}{(n+s)}\right] \end{aligned} \quad (15)$$

The generalization is proved to be valid using mathematical induction.

1. From (13) it is defined that

$$P(n''_{i_1 i_2 \dots i_m}) = P(n_{i_1 i_2 \dots i_m}) + P(s_{i_1 i_2 \dots i_m}) \approx \frac{\frac{1}{2^m}\frac{n}{m} + \frac{1}{2^m}\frac{s}{m}}{\frac{n+s}{m}} = \frac{1}{2^m}$$

   a. If $m = 1$ then it was already proved in (7)
   b. If $m = k$ then we will have

$$\begin{aligned} P(n'_{i_1 i_2 \dots i_k}) &= P(n_{i_1 i_2 \dots i_k}) + P(s_{i_1 i_2 \dots i_k}) \\ &\approx \frac{\frac{1}{2^k}\frac{n}{k}}{\frac{n+s}{k}} + \frac{\frac{1}{2^k}\frac{s}{k}}{\frac{n+s}{k}} \\ &= \frac{\frac{1}{2^k}\left(\frac{n+s}{k}\right)}{\frac{n+s}{k}} = \frac{1}{2^k} \end{aligned}$$

   c. Suppose for the case $m = k$ holds $P(n'_{i_1 i_2 \dots i_m}) = \frac{1}{2^k}$ then for $m = k+1$ we have

$$\begin{aligned} P(n'_{i_1 i_2 \dots i_m}) &= P(n'_{i_1 i_2 \dots i_k}) + P(n'_{i_1}) \\ &\approx \frac{\left(\frac{1}{2^k}\right)\left(\frac{n+s}{k}\right)}{\frac{n+s}{k}} + \frac{\left(\frac{1}{2}\right)\left(\frac{n+s}{2}\right)}{\frac{n+s}{2}} \\ &= \frac{\left(\frac{1}{2^k}+\frac{1}{2}\right)\left(\frac{(n+s)}{k}+\frac{(n+s)}{2}\right)}{\frac{(n+s)}{k}+\frac{(n+s)}{2}} = \frac{1}{2^{k+1}} \end{aligned}$$

It is proved that the condition 1 holds for $m = k+1$. So it can be concluded that the theorem point 1 holds.

2. From (14) we have

$$\begin{aligned} P(n'_{i_1 i_2 \dots i_m}) &= P(n_{i_1 i_2 \dots i_m}) + P(s_{i_1 i_2 \dots i_m}) \\ &\approx \frac{(n+s)+2^m(\delta s)}{2^m(n+s)} \\ &= \left[\frac{1}{2^m} + \frac{\delta s}{(n+s)}\right] \end{aligned}$$

   a. If $m = 1$ then it was proved by (8) that the $P(n'_i) = \left[\frac{1}{2} + \frac{\delta s}{(n+s)}\right]$ for $i \epsilon \{0,1\}$

   b. If $m = k$ then we have

$$\begin{aligned} P(n'_{i_1 i_2 \dots i_k}) &= P(n_{i_1 i_2 \dots i_k}) + P(s_{i_1 i_2 \dots i_k}) \\ &\approx \frac{\frac{1}{2^k}\cdot\frac{n}{k} + \left(\frac{1}{2^k}+\delta\right)\frac{s}{k}}{\frac{n+s}{k}} \\ &= \frac{(n+s)+2^k(\delta s)}{2^k(n+s)} \\ &= \left[\frac{1}{2^k} + \frac{\delta s}{(n+s)}\right] \end{aligned}$$

   c. Let $m = k$ holds then for $m = k+1$ we will have

$$\begin{aligned} P(n'_{i_1 i_2 \dots i_m}) &= P(n'_{i_1 i_2 \dots i_k}) + P(n'_{i_1}) \\ &\approx \left(\frac{\frac{1}{2^k}\cdot\frac{n}{k}}{\frac{n+s}{k}} + \frac{\left(\frac{1}{2^k}+\delta\right)\frac{s}{k}}{\frac{n+s}{k}}\right) + \left(\frac{\left(\frac{1}{2}n\right)}{\frac{n+s}{2}} + \frac{\left(\frac{1}{2}+\delta\right)s}{\frac{n+s}{2}}\right) \\ &= \frac{(n+s)+2^k(\delta s)}{2^k(n+s)} + \frac{(n+s)+2\delta s}{2(n+s)} \\ &\approx \left[\frac{1}{2^k} + \frac{\delta s}{(n+s)}\right] + \left[\frac{1}{2} + \frac{\delta s}{(n+s)}\right] \\ &= \left(\frac{1}{2^k} + \frac{1}{2}\right) + \frac{2\delta s}{n+s} \\ &\approx \frac{1}{2^{k+1}} + \frac{\delta s}{n+s} \end{aligned}$$

Since the value of $2\delta$ is positive and relatively small then we can consider it as another constant defined as δ. Therefore the formulation also holds for $m = k+1$. This proof also holds for the formulation (15).

Based on the general proofs above for *m*-bits pattern, then we propose the second Theorem.

**Theorem 2: *m*-bit insertion attack with random bits.**

Suppose *U* is a random sequence with *n* bits length. And *V* is a sequence of bits with length *s* to be inserted into sequence *U*. Then the effects of insertion attack on position *nm*, m > 2, *n* = 1,2,3,... with non-overlapping pattern will have two results:

1. The target sequence will have uniform distribution after the attack with probability of each pattern is $\frac{1}{2^m}$ if only if the distribution of each *m*-bit pattern in *V* is also uniform.
2. If the probability of an *m*-bit pattern in sequence *V* to be inserted is not balance, then the probability of an *m*-bit pattern to occur after the attack will bias from $\frac{1}{2^m}$ with an error $\pm\frac{\delta s}{(n+s)}$., depends on the sign of the bias in *V*.

**Corollary 2:**

The *m*-bit insertion attack on position *nm*, *n* = 1,2,3,... potential to reduce or damage the randomness property of the random sequence *U* after the attack, if the probability of each *m*-bit pattern in *V* to be inserted into *U* are not balance.

Next sub section discusses the proofs of insertion attack effect with random bits on position instead of *nm* which is notated as position on *nm+t*, *n* = 1,2,3, ... , *t* = 1,2, ..., (*m*-1).

*2) On position nm+t, n = 1,2,3, ... , t = 1,2, ..., (m-1)*

Insertion attack on position *nm+t*, *n* = 1,2,3,..., *t* = 1,2,..., (*m*-1) gave different effects on the distribution of an *m*-bits pattern in the posattack sequence. As can be seen from Fig.2 the bits before and after insertion point are forming 2 new patterns depend on the inserted bits.

For pattern *m* = 2 bits, each possible insertion attack on position odd will create 2 new patterns such that the probability of creating a new pattern on an insertion attack event is $\approx \frac{1}{2}$. Since there are 4 possible insertion events that are independent each other, then from probability theory we could say that probability of creating a new pattern on an insertion event is ½ x ¼ = 1/8. Statistically, the problem of creating new pattern can be formulated as follow:

Suppose *E* is an event of creating a new pattern by inserting a pattern *a* on an insertion attack event. Since each insertion event *E* creates two new patterns then the event *E* is defined as $E = \{G_1, G_2\}$, where $P(E = G_1) = P(E = G_2) \approx \frac{1}{2}$.

From Fig 3, there are 4 possible ways to insert a 2-bits pattern *a* between two consecutive bits. Suppose *F* is an insertion event of a certain 2-bits pattern *a* into a certain possible 2-bits pattern b. Since there are 4 possible patterns of b then we can write the event $F = \{F_1, F_2, F_3, F_4\}$. Let $F_i$ has the same probability to occur, then $P(F_i) \approx \frac{1}{4}$ for $i = 1,2,3,4$.

Events *E* and *F* are independent, then the probability of creating a new pattern *E* on insertion event *F* can be defined as $P(E \cap F) = P(E)x\,P(F) \approx \frac{1}{2}\,x\,\frac{1}{4} = \frac{1}{8}$. Fig. 3 shows all possible new patterns that will be created by inserting pattern 11 into 4 possible patterns 00, 01, 10 and 11.

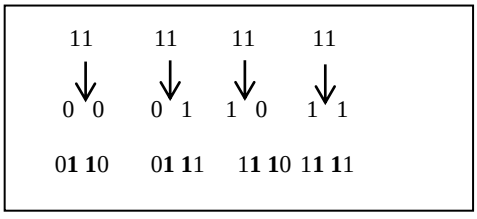


Fig. 3 All possible created patterns from inserting pattern 11

Suppose we want to count the probability of creating new pattern 11 by inserting pattern 11 on each possible insertion attack event. From Fig 3 we get $E1 = \{01,10\}$, $E2 = \{01,11\}, E3 = \{11,10\}$ *and* $E4 = \{11,11\}$, so that we found $P(E_1 = 11) = 0, P(E_2 = 11) = \frac{1}{2}, P(E_3 = 11) = \frac{1}{2}$, and $P(E_4 = 11) = 1$.

Because $E_i$ is disjoint for every *i*, then the probability of creating pattern 11 by inserting pattern 11 on each possible event, notated as $P(E_i = 11|11)$, can be determined as

$$\begin{aligned} P(E = 11|11) &= P(E_1).P(F) + P(E_2).P(F) \\ &\quad + P(E_3).P(F) + P(E_4).P(F) \\ &\approx 0 + \frac{1}{2}.\frac{1}{4} + \frac{1}{2} + \frac{1}{4} + 1.\frac{1}{4} = \frac{4}{8} \end{aligned}$$

The probability value 4/8 above is based on assumption that the insertion attack took place on all possible patterns 00, 01, 10, and 11. Or in other words the insertion attack events are balance for all possible patterns.

With the same way, we could also count the probability of creating new pattern 11 by inserting another pattern 10, 01, and 00 in all possible patterns as described below:

- $P(E = 11|10) \approx \frac{2}{8}$
- $P(E = 11|01) \approx \frac{2}{8}$
- $P(E = 11|00) \approx 0$

All the probability of creating pattern 11 by inserting all possible 2-bits patterns in four possible insertion events can be seen in Fig 4.

Pr'(11)

Insertion event

| Inserted bits | 00 | 01 | 10 | 11 | Total |
|---|---|---|---|---|---|
| 00 | 0 | 0 | 0 | 0 | 0 |
| 01 | 0 | 1 | 0 | 1 | 2 |
| 10 | 0 | 0 | 1 | 1 | 2 |
| 11 | 0 | 1 | 1 | 2 | 4 |
| Total | 0 | 2 | 2 | 4 | 8 |

Fig. 4 Distribution of creating pattern 11 for all possible insertion events

We could do the same rule to determine the occurrence probability of all other possible patterns on all possible insertion attack events as shown on Fig. 5 to Fig. 7.

Pr'(00)

Insertion events

| Inserted bits | 00 | 01 | 10 | 11 | Total |
|---|---|---|---|---|---|
| 00 | 2 | 1 | 1 | 0 | 4 |
| 01 | 1 | 1 | 0 | 0 | 2 |
| 10 | 1 | 0 | 1 | 0 | 2 |
| 11 | 0 | 0 | 0 | 0 | 0 |
| Total | 4 | 2 | 2 | 0 | 8 |

Fig. 5 Distribution of creating pattern 00 for all possible insertion events

Pr'(01)

Insertion events

| Inserted bits | 00 | 01 | 10 | 11 | Total |
|---|---|---|---|---|---|
| 00 | 0 | 0 | 1 | 1 | 2 |
| 01 | 1 | 0 | 2 | 1 | 4 |
| 10 | 0 | 0 | 0 | 0 | 0 |
| 11 | 1 | 0 | 1 | 0 | 2 |
| Total | 2 | 0 | 4 | 2 | 8 |

Fig. 6 Distribution of creating pattern 01 for all possible insertion events

Pr'(10)

Insertion event

| Inserted bits | 00 | 01 | 10 | 11 | Total |
|---|---|---|---|---|---|
| 00 | 0 | 1 | 0 | 1 | 2 |
| 01 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 | 2 | 0 | 1 | 4 |
| 11 | 1 | 1 | 0 | 0 | 2 |
| Total | 2 | 4 | 0 | 2 | 8 |

Fig. 7 Distribution of creating pattern 10 for all possible insertion events

Based on the phenomenon as is described above, then the 2-bit insertion attack on position odd will have different probability for each possible pattern on each possible event. This condition leads to a conclusion that the probability of an 2-bits pattern after the attack on position odd using NOL approach cannot be generalized. We only can limit the formulation for each case.

For example, the occurrence probability of pattern 11 by inserting pattern 11 on position odd with NOL approach can be defined as

$$P(n'_{11}) = P(n_{11}) + P(E = 11|11) \approx \frac{\frac{1}{4}\frac{n}{2} + \frac{4}{8}\frac{s}{2}}{\frac{n+s}{2}} = \frac{n+2s}{4(n+s)} = \frac{n+s+s}{4(n+s)} = \frac{1}{4} + \frac{s}{4(n+s)}$$

Meanwhile the occurrence probability of pattern 11 by inserting another pattern 10, 01, and 00 on position odd with NOL approach are defined as

- $P(n'_{11}) = P(n_{11}) + P(E = 11|10) \approx \frac{\frac{1n}{42}+\frac{2s}{82}}{\frac{n+s}{2}} = \frac{n+s}{4(n+s)} = \frac{1}{4}$
- $P(n'_{10}) = P(n_{10}) + P(E = 11|00) \approx \frac{\frac{1n}{42}+\frac{2s}{82}}{\frac{n+s}{2}} = \frac{n+s}{4(n+s)} = \frac{1}{4}$
- $P(n'_{11}) = P(n_{11}) + P(E = 11|00) \approx \frac{\frac{1n}{42}+0.\frac{s}{2}}{\frac{n+s}{2}} = \frac{n}{4(n+s)} = \frac{n+s-s}{4(n+s)} = \frac{1}{4} - \frac{s}{4(n+s)}$

From the examples above we see the occurrence probability of pattern 11 by inserting pattern bit 10 and 01 where $P(E = 11|01) = P(E = 11|10) = 2/8$ is the same i.e 1/4, meanwhile the occurrence probability of pattern 11 to occur by inserting pattern 11, where the $P(E = 11|11) = 4/8$, is $\frac{1}{4} + \frac{s}{4(n+s)}$ contradictive with the case of inserting pattern 00 with $P(E = 11|00) = 0$ that produced the probability $\frac{1}{4} - \frac{s}{4(n+s)}$ for pattern 11 to occur after the attack.

From the results above, the occurrence probability of 2-bits pattern after insertion attack on position odd with NOL approach is vary based on the pattern of inserted bits and also the pattern of bits where the insertion took place. Generally, the occurrence probability of 2-bits pattern after the attack is really depend on $P(E)$.

This conditions also hold for insertion attack with $m > 2$, on position $nm+t$, $n = 1,2,3,...$; $t = 1,2,..., (m-1)$. For example for $m = 3$, we can find two new patterns created on each insertion attack event, where the total number of new patterns space is 16 that taken from 2 (two) possible positions $3n+1$ and $3n+2$ for $n = 1,2,3,...$ The example of occurrence probability of pattern 000 after inserted 3-bits pattern on position $3n+1$ and $3n+2$ can be seen in Table 3.

TABLE III
FREQUENCY OF PATTERN 000 CREATED AFTER INSERTED 3-BITS ON POSITION $3N+1$ AND $3N+2$.

| Pos'n $3n+1$ 000 ↓ | Pattern created | Freq of 000 | Pos'n $3n+2$ 000 ↓ | Pattern created | Freq of 000 |
|---|---|---|---|---|---|
| 0 00 | 000 000 | 2 | 00 0 | 000 000 | 2 |
| 0 01 | 000 001 | 1 | 00 1 | 000 001 | 1 |
| 0 10 | 000 010 | 1 | 01 0 | 010 000 | 1 |
| 0 11 | 000 011 | 1 | 01 1 | 010 001 | |
| 1 00 | 100 000 | 1 | 10 0 | 100 000 | 1 |
| 1 01 | 100 001 | | 10 1 | 100 001 | |
| 1 10 | 100 010 | | 11 0 | 110 000 | 1 |
| 1 11 | 100 011 | | 11 1 | 110 001 | |
| 8 | 16 | 6 | 8 | 16 | 6 |

Based on the two cases on $m = 2$ and $m = 3$ the occurance probability of an $m$-bits pattern after the attack is different for each pattern depends on the value of $P(E)$. Therefore in this case, the occurrence probability of an $m$-bits pattern after insertion attack with random bits on position $nm+t$, $n = 1,2,3,...$; $t = 1,2,..., (m-1)$, depends on the probability of creating new $m$-bits patterns based on all $2^m$ possible insertion of $m$-bits pattern in all $2^m$ possible insertion events.

By assuming that $P(E)$ is the probability of creating an $m$-bits pattern $n'_{i_1 i_2 \dots i_m}$ by inserting a certain $m$-bits pattern $s_{i_1 i_2 \dots i_m}$ on all possible insertion events $h_{i_1 i_2 \dots i_m}$, on position $nm+t$, $n = 1,2,3,...$ , $t = 1,2,3,...,(m-1)$, the occurrence probability of an $m$-bits pattern after the attack with NOL approach can be defined as:

$$P(n'_{i_1 i_2 \dots i_m}) = P(n_{i_1 i_2 \dots i_m}) + P(E)$$

$$= \frac{P(n_{i_1 i_2 \dots i_m})\frac{n}{m} + P(n'_{i_1 i_2 \dots i_m} | s_{i_1 i_2 \dots i_m})\frac{s}{m}}{(n+s)/m}$$

where $i_1 i_2 \dots i_m \in \{0,1\}^m$. (16)

The next subsection will discuss about the insertion attack effects with OL pattern approach.

*D. M-bit insertion attack with OL pattern approach*

The insertion attack with OL approach also indicates the same problem as happened on the insertion attack with NOL approach on position $nm+t$, $n$ = 1,2,3,...; $t$ = 1,2,..., ($m$-1).

The different is only on the number of patterns created on each insertion attack event. For example on 2-bits insertion attack with OL approach, there are 3 new patterns created for each insertion attack event, instead of 2 patterns as in NOL approach on position odd, so that the total number of new patterns is 12 patterns as is shown in Fig. 8.

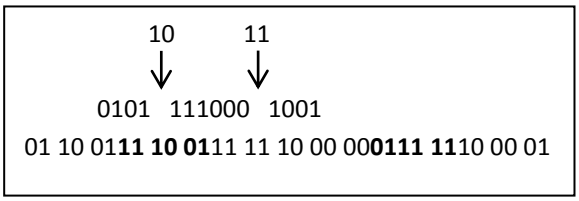


Fig.8 2-bits insertion attack with OL on even position

The patterns that are created after the insertion attack with OL approach are depending on the pattern of bits inserted and the value of bits before and after the insertion point.

Based on probability principles, the probability of creating a pattern 00 by inserting pattern 01 in all 4 possible events, notated as $P(E = 00|01)$ can be counted as

$$P(E = 00|01) = P(E_1).P(F1) + P(E_2).P(F_2) + P(E_3).P(F_3) + P(E_4).P(F_4)$$
$$\approx 1.\frac{1}{4} + \frac{1}{3}.\frac{1}{4} + \frac{1}{3}.\frac{1}{4} + 0 = \frac{5}{12}$$

With the same rules, the probability of creating pattern 00 by inserting all possible patterns 01, 10, 11, and 00, in 4 possible ways, can be determined as is shown in Fig 9.

Pr' (00)

Inserted bits \ Insertion event

| | 00 | 01 | 10 | 11 | Σ |
|---|---|---|---|---|---|
| 00 | 3 | 1 | 1 | 0 | 5 |
| 01 | 2 | 1 | 0 | 0 | 3 |
| 10 | 2 | 0 | 1 | 0 | 3 |
| 11 | 1 | 0 | 0 | 0 | 0 |
| Σ | 8 | 2 | 2 | 0 | 12 |

Fig. 9 Distribution of new patterns 00 for 2-bits insertion with OL approach

Based on Fig 9, the occurrence probability of pattern 00 in postattack sequence after inserting pattern 01 with OL approach can be defined as

$$P(n'_{00}) = P(n_{00}) + P(E = 00|01)$$
$$\approx \frac{\frac{1}{4}n + \frac{5}{12}s}{n+s} = \frac{3n+5s}{12(n+s)}$$
$$= \frac{3n+3s+2s}{12(n+s)} = \frac{3(n+s)+2s}{12(n+s)} = \frac{1}{4} + \frac{s}{6(n+s)}$$

Meanwhile the occurrence probability of pattern 00, by inserting pattern 10, 01, and 00, with OL approach can be defined as

- $P(n'_{00}) = P(n_{00}) + P(E = 00|10)$
$\approx \frac{\frac{1}{4}n + \frac{3}{12}s}{n+s} = \frac{3n+3s}{12(n+s)} = \frac{1}{4}$

- $P(n'_{00}) = P(n_{00}) + P(E = 00|01)$
$\approx \frac{\frac{1}{4}n + \frac{3}{12}s}{n+s} = \frac{3n+3s}{12(n+s)} = \frac{1}{4}$

- $P(n'_{00}) = P(n_{00}) + P(E = 00|00)$
$\approx \frac{\frac{1}{4}n + 0}{n+s} = \frac{n+s-s}{4(n+s)} = \frac{1}{4} - \frac{s}{4(n+s)}$

From the results above, the occurrence probability of pattern 00 in postattack sequence is vary and really depends on the $P(E)$. With the same way we could count the occurrence probability of other pattern 01, 10, and 11 by inserting all possible patterns 00, 01, 10, and 11 in all 4 possible ways of insertion attack as are shown in Fig 10 to Fig 12.

Pr'(01)

Inserted bits \ Insertion events

| | 00 | 01 | 10 | 11 | Total |
|---|---|---|---|---|---|
| 00 | 0 | 1 | 1 | 1 | 3 |
| 01 | 1 | 1 | 2 | 1 | 5 |
| 10 | 0 | 1 | 0 | 0 | 1 |
| 11 | 0 | 1 | 0 | 2 | 3 |
| Total | 1 | 4 | 3 | 4 | 12 |

Fig. 10 Distribution of patterns 01 for 2-bits insertion with OL approach

Pr'(10)

Inserted bits \ Insertion events

| | 00 | 01 | 10 | 11 | Total |
|---|---|---|---|---|---|
| 00 | 0 | 1 | 1 | 1 | 3 |
| 01 | 0 | 0 | 1 | 0 | 1 |
| 10 | 1 | 2 | 1 | 1 | 5 |
| 11 | 1 | 1 | 1 | 0 | 3 |
| Total | 2 | 4 | 4 | 2 | 12 |

Fig. 11 Distribution of patterns 10 for 2-bits insertion with OL approach

Pr'(11)

Inserted bits \ Insertion events

| | 00 | 01 | 10 | 11 | Total |
|---|---|---|---|---|---|
| 00 | 0 | 0 | 0 | 1 | 1 |
| 01 | 0 | 1 | 0 | 2 | 3 |
| 10 | 0 | 0 | 1 | 2 | 3 |
| 11 | 0 | 1 | 1 | 3 | 5 |
| Total | 0 | 2 | 2 | 8 | 12 |

Fig.12. Distribution of patterns 11 for 2-bits insertion with OL approach

The same condition also holds for $m > 2$. Based on this facts, the occurrence probability of an *m*-bits pattern after the insertion attack with OL approach cannot be generalized but can only limited on probability of *P(E)* such as formulated in (17).

For a sequence with length *n*-bits, the total cyclic overlapping pattern is *n* pattern, so that the occurrence probability of a pattern $n'_{i_1 i_2 \dots i_m}$ in postattack sequence can be defined as

$$P(n'_{i_1 i_2 \dots i_m}) = P(n_{i_1 i_2 \dots i_m}) + P(E) = \frac{P(n_{i_1 i_2 \dots i_m})n + P(n'_{i_1 i_2 \dots i_m} | s_{i_1 i_2 \dots i_m})s}{(n+s)}$$
where $i_1 i_2 \dots i_m \in \{0,1\}^m$. (17)

Based on the proofs above, the insertion attack effect is vary depends on how the attack is conducted. The attack potentialy changes the distribution of the probability of each pattern to occur in the postattack sequence that could cause the damage of the randomness property of the target sequence. Tabel IV showed the attack effects based on the occurance probability of an *m*-bits pattern in the postattack sequence.

The next section presents the effects of the insertion attack with extreme bits (i.e. the same bits occured consecutively) that intuitively must destroy the randomness property of the target sequence.

TABLE IV
THE INSERTION ATTACK EFFECTS WITH RANDOM BITS

| | Position | Assumption | $P(n'_{i_1 i_2 \dots i_m})$ |
|---|---|---|---|
| NOL | *mn* *n*= 1,2,3, … | $P(s_{i_1 i_2 \dots i_m}) = \frac{1}{2^m}$ | $\frac{1}{2^m}$ |
| | | $P(s_{i_1 i_2 \dots i_m}) > \frac{1}{2^m}$ | $\frac{1}{2^m} + \frac{m\delta}{(n+s)}$ |
| | | $P(s_{i_1 i_2 \dots i_m}) < \frac{1}{2^m}$ | $\frac{1}{2^m} - \frac{m\delta}{n+s}$ |
| | *mn+t* *t* = ,1,2,...,m-1 *n*= 1,2,3,... | *P(E)* depend on bits before and after insertion point | $\frac{n_{i_1 i_2 \dots i_m} + P(E)}{(n+s)/m}$ |
| OL | - | *P(E)* depend on bits before and after insertion point | $\frac{n_{i_1 i_2 \dots i_m} + P(E)}{(n+s)}$ |

## IV. INSERTION ATTACK EFFECTS WITH EXTREME BITS

The insertion attack with extreme bits is conducted by inserting an extreme patterns i.e. 1111...1 or 0000...0 or 10101010...10 or other forms into the target sequence *U*. Intuitively the sequence should be not random after the attack. Because the *s*-bits pattern in *V* that will be inserted in sequence *U* are constant values where $P(S_{i_1,i_2,\dots,i_m}) = 1$ so that the expected value of $S_{i_1,i_2,\dots,i_m}$ will be $P(S_{i_1,i_2,\dots,i_m}).s = 1.s = s$.

On the attack with extreme bits also holds the same cases where the attack effects can only be generalized when the attack took place on position *nm* with NOL pattern approach. The occurrence probability of bit 1 (or 0) after 1-bit pattern attack is indicated in (18).

$$P(n'_{i_1}) = P(n_{i_1}) + P(s_{i_1}) \approx \frac{\frac{1}{2}n + 1.s}{n+s} = \frac{n+2s}{2(n+s)} = \frac{(n+s)+s}{2(n+s)} = \frac{1}{2} + \frac{s}{2(n+s)} \quad (18)$$

Noted, $n'_{i_1}$ is the bit inserted. In the case that the extreme bits inserted is 1111...1 then the occurence of the complement bit 0 will remain $\frac{n}{2}$ as in the preattack sequence *U*. The occurance probability of bit 0 in the postattack sequence *U+V* is defined as

$$\frac{\frac{1}{2}(n)+0}{n+s} = \frac{n}{2(n+s)} \quad (19)$$

For 2-bits extreme pattern attack for e.g., 10101010...10, the occurrence probability of pattern 10 in the postattack sequence can be determined as

$$P(n'_{i_1 i_2}) = \frac{1}{4} + \frac{3s}{4(n+s)} \quad (20)$$

where $n'_{i_1 i_2}$ is the inserted 2-bits pattern. Meanwhile the occurance probability of other 2-bits patterns notated as $P(\sim n'_{i_1 i_2})$ after the attack are remain the same as in the preattack sequence which is defined in (21).

$$P(\sim n'_{i_1 i_2}) = \frac{\frac{1}{4}\left(\frac{n}{2}\right)+0}{\frac{(n+s)}{2}} = \frac{n}{4(n+s)} \quad (21)$$

With the same way for 3-bits extreme pattern attack, the probability of $n'_{i_1 i_2, i_3}$ and $\sim n'_{i_1 i_2, i_3}$ respectively are defined in (22) and (23).

$$P(n'_{i_1 i_2, i_3}) = \frac{1}{8} + \frac{7s}{8(n+s)} \quad (22)$$

$$P(\sim n'_{i_1 i_2, i_3}) = \frac{n}{8(n+s)} \quad (23)$$

From the formulation in (18) to (23) it is showed that the attack effects can be generalized for *m*-bits pattern i.e.:

$$P(n'_{i_1 i_2, \dots, i_m}) = \frac{\frac{1}{2^m}\left(\frac{n}{m}\right) + \frac{s}{m}}{\frac{n+s}{m}} = \frac{n + 2^m s}{2^m(n+s)} = \frac{1}{2^m} + \frac{(2^m - 1)s}{2^m(n+s)} \quad (24)$$

$$P(\sim n'_{i_1 i_2, \dots, i_m}) = \frac{\frac{1}{2^m}\left(\frac{n}{m}\right)+0}{\frac{n+s}{m}} = \frac{n}{2^m(n+s)} \quad (25)$$

where the $n'_{i_1 i_2, \dots, i_m}$ is the inserted *m*-bits pattern and $\sim n'_{i_1 i_2, \dots, i_m}$ is another *m*-bits patterns.

Using mathematical induction, it can be proved that the generalization holds. For the case, we propose Theorem 3 about insertion attack effects with extreme bits with NOL approach on position *nm*.

**Theorem 3** (Insertion attack effects with extreme bits on position *nm* using NOL approach)

Suppose a sequence *U* with periodic *p*, $U = u_1 u_2 u_3 \dots u_p$, and let *U* is random and uniform so that $P(n_{i_1 i_2 \dots i_m}) = \frac{1}{2^m}$ for $i_1 i_2 \dots i_m \in \{0,1\}$, and $m \geq 1$. Let a sequence *V* with length *s*-bits contains extreme *m*-bits patterns will be

inserted in *U* on position *nm*, for *n* = 1,2,3, … using NOL approach. Then after the attack:

1. The occurrence probability of inserted *m*-bits pattern will be $P(n_{i_1 i_2 \dots i_m}) \approx \frac{1}{2^m} + \frac{(2^m-1)s}{2^m(n+s)}$ with bias $\frac{(2^m-1)s}{2^m(n+s)}$ from $\frac{1}{2^m}$.
2. The occurrence probability of other *m*-bits patterns will be $P(\sim n_{i_1 i_2 \dots i_m}) \approx \frac{n}{2^m(n+s)}$ which remains the same as in the preattack sequence.

The bias of occurrence probability of an *m*-bits pattern for the attack with NOL approach on position $nm + t, n = 1,2,\dots;\ t = 1,2,\dots,(m-1)$ or with OL approach cannot be generalized, because the probability of new patterns created after the attack depends on the bits before and after the insertion point notated as $P(E)$.

For example, on 2-bits pattern 11 attack, $P(E)$ is limited on the insertion of pattern 11 on four possible patterns 00, 01, 10, and 11. As is defined on Fig 4 for NOL approach on position *mn+t* and Fig 12 for OL approach, the occurrence probability of pattern 11 are $\frac{4}{8}$ and $\frac{8}{12}$ respectively, which is higher than other patterns. So it proved that the postattack sequence will be no longer balance that is potential to loose the randomness property. Entirely, the effects of insertion attack with extreme bits can be seen in Table V.

TABLE V
THE INSERTION ATTACK EFFECTS WITH EXTREME BITS

| | POSITION | ASSUMPTION | $P(n'_{i_1 i_2 \dots i_m})$ |
|---|---|---|---|
| NOL | *MN*<br>*N*= 1,2,3, … | $P(s_{i_1 i_2 \dots i_m}) = 1$ | $\frac{1}{2^m} + \frac{(2^m-1)s}{2^m(n+s)}$ |
| | *MN+T*<br>*T* = 1,2,…,M-1<br>*N*= 1,2,3, … | $P(s_{i_1 i_2 \dots i_m}) = 1$ | $\frac{n_{i_1 i_2 \dots i_m} + P(E)}{(n+s)/m}$ |
| OL | - | $P(s_{i_1 i_2 \dots i_m}) = 1$ | $\frac{n_{i_1 i_2 \dots i_m} + P(E)}{(n+s)}$ |

From Table V it can be seen that the insertion attack with extreme bits positively causes the bias for the occurence probability of *m*-bits pattern after the attack. And as the conclusion of the insertion attack with NOL approach on position *nm+t* and OL approach, we proposed the last theorem.

**Theorem 4** (Insertion attack effects with NOL approach on position $nm + t, n = 1,2,\dots;\ t = 1,2,\dots,(m-1)$ and OL approach).

Suppose a sequence *U* with periodic *p*, $U = u_1 u_2 u_3 \dots u_p$, and let *U* is random and uniform so that $P(n_{i_1 i_2 \dots i_m}) = \frac{1}{2^m}$ for $i_1 i_2 \dots i_m \in \{0,1\}$, and $m \geq 1$. Let a sequence *V* with length *s*-bit will be inserted in sequence *U* then:

1. The occurrence probability of *m*-bit after the attack on position $nm + t, n = 1,2,\dots;\ t = 1,2,\dots,(m-1)$ with NOL approach is :

$$P(n'_{i_1 i_2 \dots i_m}) = P(n_{i_1 i_2 \dots i_m}) + P(E)$$

$$= \frac{P(n_{i_1 i_2 \dots i_m})n + P(n'_{i_1 i_2 \dots i_m} | s_{i_1 i_2 \dots i_m}) \frac{s}{m}}{(n+s)/m}$$

where $i_1 i_2 \dots i_m \in \{0,1\}^m$.

2. The occurrence probability of *m*-bit after the attack with OL approach is :

$$P(n'_{i_1 i_2 \dots i_m}) = P(n_{i_1 i_2 \dots i_m}) + P(E)$$

$$= \frac{P(n_{i_1 i_2 \dots i_m})n + P(n'_{i_1 i_2 \dots i_m} | s_{i_1 i_2 \dots i_m})s}{(n+s)}$$

where $i_1 i_2 \dots i_m \in \{0,1\}^m$.

3. The conditional probability in point 1 and point 2 each holds for insertion attack with random bits or with extreme bits.

Theorem 3 and 4 have proved that the insertion attack with extreme bits in all possible ways will postively cause the postattack sequences no longer uniform so that is potential to loose the randomness property. Meanwhile Theorem 1, 2 and 4 have proved that the insertion attack with random bits is potential to reduce the randomness property of the target sequence if the insertion attack is not balance. Otherwise, when the insertion attack is balance, the postattack sequences remains uniform that potential to maintain the randomness property.

In practice, this kind of attack is possible to conduct on the output sequence of an RNG (or a PRNG) using software approach or hardware approach (see Fig.13).

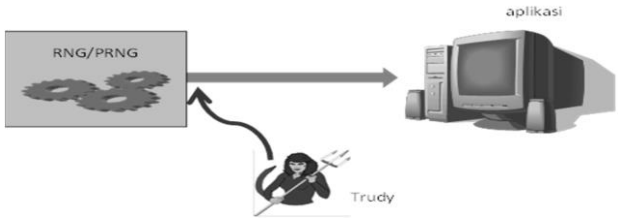

Fig. 13 Insertion attack on RNG/PRNG

For example the attack can be conducted by injecting a Trojan into a security system that is designed to attack the output sequence of an RNG (or a PRNG) inside the system while producing the random sequences. The goal is to make the output sequence bias. From the empirical studies in previous reseach and also the proofs discussed in this paper, the insertion attack potentially reduces or damages the randomness property of the target sequence.

Some other researches showed that this kind of attack practically fisible to implement. For eample in [16], Markettos and More proved that the randomness property of an RNG in EMV payment card are destroyed by injecting the signal into the EMV card in such a way. By little modification on the device, this attack significantly could reduce the possibility of key spaces from $2^{32}$ into only $2^8$ that make it possible to masquarade the transactions. Another related research in [17], Becker, *et.al.*, showed that Dopant-based Trojan they proposed could compromised the RNG used in processor Ivy Bridge from Intel so that the security of random sequences produced is reduced from $2^{128}$ into $2^n$ for $n < 128$ where *n* is chosen by the attacker.

This two researches proved that the insertion attacked proposed in this paper is also possible to conduct in practice.

## V. Conclusion

The insertion attack with random bits or with extreme bits are potential to change the distribution of occurance probability of an *m*-bits pattern in postattack sequence. The measurements of the effects based on the occurrence probability of an *m*-bits pattern in the postattack sequence gave four fundamental results:

1. The balance of insertion attacks will keep the uniformity of each *m*-bit pattern with probability $1/2^m$ in the postattack sequence that could potentially keep the randomness property of the sequence;
2. If the attack is not balance, it will cause the distribution of *m*-bit patterns is not uniform with bias: $\pm \frac{m\delta}{n+s}$ for insertion attack with random bits and $+\frac{(2^m-1)}{2^m(n+s)}$ on insertion attack with extreme bits. The generalization of bias above only holds for the attacks on position *nm* with NOL approach.
3. Insertion attack with NOL approach on position instead of *nm* and with OL approach cause the occurance probability of each *m*-bits pattern conditionally depends on the bits before and after the attack point. The conditional probability *P(E)* is vary for each possible insertion event and potential to cause the loss of randomness property of the target sequence. The bias occured can not be generalized.
4. In the case of attack with extreme bits, the occurence probability of the inserted *m*-bits pattern will significantly higher than other *m*-bits patterns that causes the postattack sequence to lose the randomness property.

From theoretical proofs above, we could conclude that the insertion attacks might cause the loss of randomness property of a random (or pseudorandom) sequence that should be considered carefully, especially in cryptographic applications or other applications that need a random sequence as a critical input.

Finally, it is important to find a way how to detect that such attack exists in the security system and how to anticipate the attack so that the system could manage to deliver its security services properly.

## References


[1] B. Schneier, (1996), Applied Cryptography, 2nd ed., John Wiley & Son, Inc., USA.

[2] L. Dorendorf, Z. Gutterman, and B. Pinkas, (2007), Cryptanalysis of the random number generator of the windows operating system. *ACM Trans. Inf. Syst. Secur*., 13(1):1–32, 2009.

[3] L. Bello, (2008), OpenSSL – predictable random number generator, *Debian security advisory* 1571-1, https://www.debian.org/security/2008/dsa-1571.

[4] B. Sanguinetti, A. Martin, H Zbinden, and N. Gisin, (2014), Quantum random number generator on mobile phone, *J. American Physical Society*, Physical review X 4, 031056 (2014), http://journals.aps.org/prx/pdf/10.1103/PhysRevX.4.031056.

[5] Bitcoin, (11 August 2013), Android Security Vulnerability, https://bitcoin.org/en/alert/2013-08-11-android.

[6] S. Indarjani, *1-bit Insertion Attack effect on Randomness Tplerance of Random Bit Sequences*, in *Proc. Konferensi Nasional Sistem dan Informatika* 2008; Bali, , KNS&I08-036, pp. 203-208, 2008

[7] S. Indarjani and B. Widjaja, "Measuring the insertion attack effect on randomness property of AES-based PRNG, IPCSIT vol.40 2012, pp 118-122, 2012

[8] S. Indarjani, A. Nugraha, G. Supriyatno, I.M.M. Astawa., (2014), Insertion Attack effects on some PRNGs Based on NIST Randomness Tests Tool : Case study on ANSI-X9.17, ANSIX9.31, Dragon and Rabbit Algorithms, *IEEE Proceeding on 2014 International Conference of Computer, Control, Informatics and Its Applications* (IC3INA 2014), 21-23 Oktober 2014, pp 181-186.

[9] D. Wulandari, A. Kumala, S. Nugroho, S. Indarjani, Comparison the insertion attack effects on randomness property of Dragon and Rabbit stream cipher," *IEEE Proceeding on 2013 International Conference of Computer, Control, Informatics and Its Applications* (IC3INA 2013), pp.213,218, 19-21 Nov. 2013.

[10] K. Inayah, B.E. Sukmono, R. Purwoko, S. Indarjani, Insertion attack effects on standard PRNGs ANSI X9.17 and ANSI X9.31 based on statistical distance tests and entropy difference tests, *IEEE Proceeding on 2013 International Conference of Computer, Control, Informatics and Its Applications* (IC3INA 2013), pp.219,224, 19-21 Nov. 2013.

[11] A.J. Menezes, P.C. van Oorschotet, and S.A. Vanstone, S.A., Handbook of Applied Cryptography, CRC Press LLC, USA, 1997

[12] Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., and Vo, S. (2010), NIST Special Publication 800-22 Revision 1a: A StatisticalTest Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST, US Department of Commerce, USA.

[13] G. Marsaglia, *"DIEHARD : A Battery of Tests of Randomness"*, 1996, available in: http://stat.fsu.edu/~geo/diehard.html, last accesed September 20, 2014 at 08.10.

[14] Van Tilborg, H.C.A., 2005, Encyclopedia of Cryptography and Security, Springer Science+Business Media, Inc., USA.

[15] J. Hoffstein, J. Pipher, J, and J.H. Silverman, An Introduction to Mathematical Cryptography, Springer Science+Business Media, LLC, New York, 2008.

[16] A.T. Markettos, and S.W. Moore, (2009), The Frequency Injection Attacks on Ring-Oscillator-Based True Random Number Generator, in *Proc. of the 11th International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 317 – 331, Springer-Verlag.

[17] Becker, G.T., Regazzoni, F., Paar, C. and Burlesson, W.P., (2013), Stealthy Dopant-level Hardware Trojan, in *CHES'13 Proceedings of the 15th international conference on Cryptographic Hardware and Embedded Systems*, published by Springer-Verlag, Berlin, 2013, pp. 197-214.