The Index j in RC4 is not Pseudo-random

Abstract. In this paper we provide several theoretical evidences that the pseudo-random index j of RC4 is indeed not pseudo-random.
First we show that in long term $\Pr(j = i + 1) = \frac{1}{N} - \frac{1}{N^2}$, instead of the random association $\frac{1}{N}$ and this happens for the non-existence of the condition $S[i] = 1$ and $j = i + 1$
that is mandatory for the non-existence of the Finney cycle. Further we also identify several results on non-existence of certain sequences of j.

Keywords: RC4, Non-randomness, Pseudo-random Index, Stream Cipher, Cryptography.

1 Introduction

As we all know, there are many results related to non-randomness of RC4 that received the attention in flagship level cryptology conferences and journals (see for example [3–5] and the references therein). Even after intense research for more than three decades on a few lines of RC4 algorithm, we are still amazed with new discoveries in this area of research.

As we are presenting a short note, we assume that the reader is aware of RC4 algorithm. Still let us present the algorithm briefly.

In RC4, there is a $N = 256$ length array of 8-bit integers 0 to $N - 1$, that works as a permutation. There is also an l length array of bytes K, where l may vary from 5 to 32, depending on the key length. There are also two bytes i, j, where i is the deterministic index that increases by 1 in each step and j is updated in a manner so that it behaves pseudo-randomly.

The Key Scheduling Algorithm (KSA) of RC4 is as follows:

```
– $j = 0$; for $i = 0$ to $N - 1$: $S[i] = i$;
– for $i = 0$ to $N - 1$:
      $j = j + S[i] + K[i \bmod l]$; swap$(S[i], S[j])$;
```

Next the pseudo-random bytes z are generated during the Pseudo Random Generator Algorithm (PRGA) as follows:

```
– $i = j = 0$;
– for $i = 0$ to $N - 1$:
      $i = i + 1$; $j = j + S[i]$; swap$(S[i], S[j])$; $z = S[S[i] + S[j]]$;
```

Note that all the additions here are modulo N.

2 Non-Randomness due to non-existence of Finney cycle

While there is long term suspicion that there could be problems with the psudo-randomness of j, till very recently it could not be observed or reported. In fact, in [4, Section 3.4], non-randomness of j has been studied for initial rounds and it has been commented that the distribution of j is almost uniform for higher rounds. Thus, to date, no long term pseudo-randomness of the index j has been reported.

It has been observed by Finney [1] that if $S[i] = 1$ and $j = i + 1$, then RC4 lands into a short cycle of length $N(N - 1)$.

Fortunately (or knowing this very well), the design of RC4 by Rivest considers the initialization of RC4 PRGA as $i = j = 0$. Thus, during RC4 PGRA, the Finney cycle cannot occur, i.e., if $\Pr(S[i] = 1)$, then $\Pr(j = i + 1) = 0$. This provides the non-randomness in j.

Theorem 1. *During RC4 PRGA,* $\Pr(j = i+1) = \frac{1}{N} - \frac{1}{N^2}$*, under certain usual assumptions.*

Proof. We have

$$\begin{aligned}
&\Pr(j = i + 1) = \Pr(j = i + 1, S[i] = 1) + \Pr(j = i + 1, S[i] \neq 1)\\
&= \ 0 + \Pr(j = i + 1 | S[i] \neq 1) \cdot \Pr(S[i] \neq 1)\\
&= \ 1\frac{}{N \cdot (1 - \frac{1}{N}) = \frac{1}{N} - \frac{1}{N^2}.}
\end{aligned}$$

Here we consider $\Pr(j = i + 1 | S[i] \neq 1) = \frac{1}{N}$ under usual randomness assumption (it has been checked by experiments too). Further, considering S as a random permutation, we get $\Pr(S[i] \neq 1) = 1 - \frac{1}{N}$. □

In fact, one can sharpen this result slightly by using Glimpse theorem as follows. Though it happens generally once out of N rounds during the PRGA.

Corollary 1. *During RC4 PRGA,* $\Pr(j = i + 1 | i = z + 1) = \frac{1}{N} - \frac{2}{N^2} + \frac{1}{N^3}$.

Proof. We refer to Glimpse theorem [2] that says, $\Pr(S[j] = i - z) = \frac{2}{N} - \frac{1}{N^2}$ after the swap of $S[i]$ and $S[j]$. Consider the situation when $S[i] = 1$ before the swap. That means $S[j] = 1$ after the swap.

Thus, $\Pr(S[i] = 1 | i = z + 1) = \frac{2}{N} - \frac{1}{N^2}$. Hence, we have the following:

$$
\begin{aligned}
&\Pr(j = i + 1 | i = z + 1) = \Pr(j = i + 1, S[i] = 1 | i = z + 1)\\
&+ \Pr(j = i + 1, S[i] \neq 1 | i = z + 1)\\
&= \ 0\\
&+ \Pr(j = i + 1 | S[i] \neq 1, i = z + 1)\\
&\cdot \Pr(S[i] \neq 1 | i = z + 1)\\
&= \ \frac{1}{N \cdot (1 - \frac{2}{N} + \frac{1}{N^2}) = \frac{1}{N} - \frac{2}{N^2} + \frac{1}{N^3}.}
\end{aligned}
$$

We consider the usual assumptions as in Theorem 1. □

Since we make a few assumptions, it is important to validate the results and the experimental data indeed supports the theoretical claims mentioned above.

3 Non-existent sequences of j over several rounds

Theorem 2. *During RC4 PRGA, in* 3 *consecutive rounds (*r*,* $r+1$ *and* $r+2$*),* j *cannot take* 3 *consecutive integer values. In other words, there is no* r *such that* $j_{r+2} = j_{r+1} + 1 = j_r + 2$.

Proof. Let us first consider a situation where j has increased by 1 from round r to round $r + 1$.

So $j_r + 1 = j_{r+1}$, which implies $S_r[i_r + 1] = S_{r+1}[j_{r+1}] = 1$.

For terminology, we have considered $S_r[k]$ as the value of the array at k-th index after the swap is done in round r.

In RC4 PRGA, a Finney cycle cannot happen. Hence, i_{r+1} cannot take the value of $(j_{r+1} - 1)$. Hence $S_{r+1}[i_{r+1} + 1]$ cannot be 1.

Thus it would not be possible to have $j_{r+2} = j_{r+1} + 1$. Hence the proof. □

As a corollary to the above theorem, we can prove the following probabilistic result.

Corollary 2. *During RC4 PRGA,* $\Pr(j_{r+2} = j_r + 2) = \frac{1}{N} - \frac{1}{N^2}$*, under certain usual assumptions.*

Proof. We have

$\Pr(j_{r+2} = j_r + 2) = \Pr(j_{r+2} = j_r + 2, j_{r+1} = j_r + 1) + \Pr(j_{r+2} = j_r + 2, j_{r+1} \neq j_r + 1)$
$= 0 + \Pr(j_{r+2} = j_r + 2 | j_{r+1} \neq j_r + 1) \cdot \Pr(j_{r+1} \neq j_r + 1)$
$= \frac{1}{N} \cdot (1 - \frac{1}{N})$
$= \frac{1}{N}$ - $\frac{1}{N^2}$.

Here we consider $\Pr(j_{r+2} = j_r + 2 | j_{r+1} \neq j_r + 1) = \frac{1}{N}$ under usual randomness assumption.

Further, considering S as a random permutation, we get $\Pr(j_{r+1} \neq j_r + 1) = 1 - \frac{1}{N}$. □

Theorem 3. *In at most three consecutive rounds* (r, $r+1$ *and* $r+2$)*, the value of* j *can remain constant* ($j_r = j_{r+1} = j_{r+2}$) *or in other words there cannot exist any* r *for which* ($j_r = j_{r+1} = j_{r+2} = j_{r+3}$).

Proof. Let us denote the difference between the j values of two consecutive rounds as d, i.e., $d_{r+1} = (j_{r+1} - j_r)$.

Clearly, $S_r[i_r + 1] = S_{r+1}[j_{r+1}] = d_{r+1}$.

We prove the claim now by contradiction.

Let us assume that it is possible to find an r and a permutation S such that $j_r = j_{r+1} = j_{r+2} = j_{r+3}$

So, $d_{r+1} = d_{r+2} = d_{r+3} = 0$, which implies -

1. $S_{r+1}[j_{r+1}] = S_{r+1}[i_{r+1} + 1] = 0$ and
2. $S_{r+2}[i_{r+2} + 1] = 0$

From 1, one can derive that

$j_{r+1} = i_{r+1} + 1 = i_{r+2}$

Since $S_{r+1}[i_{r+1} + 1] = 0$, the j value will not change from $(r+1)$ to $(r+2)$ round.

Therefore, $j_{r+2} = j_{r+1} = i_{r+2}$ and $S_{r+2}[j_{r+2}] = S_{r+2}[i_{r+2}] = 0$

Hence, $S_{r+2}[i_{r+2} + 1]$ cannot be 0 and that contradicts equation 2. Thus we prove the lemma by contradiction. □

Corollary 3. *If* $(j_r = j_{r+1} = j_{r+2})$ *then* $i_{r+2} = j_{r+2}$ *and* $S_{r+1}[j_{r+1}] = S_{r+2}[i_{r+2}] = S_{r+2}[j_{r+2}] = 0$

Proof. The result has already been proved in intermediate steps of the previous theorem.

Corollary 4. *In two consecutive rounds* $(r$ *and* $r+1)$*, if the value of* j *remains constant* $($*i.e.,* $j_r = j_{r+1})$ *then* $S_{r+1}[j_{r+1}]$ *must be* 0.

Proof. Using the same notation as used in Corollary 3 we observe that $d_{r+1} = (j_{r+1} - j_r) = 0$

Therefore, $S_{r+1}[j_{r+1}] = S_r[i_r + 1] = d_{r+1} = 0$. Hence the proof. □

Corollary 5. *Once a value of* j *gets repeated in three consecutive rounds* $(r$*,* $r+1$ *and* $r+2)$*, no value can immediately be repeated in subsequent two rounds* $($*for* $N > 2)$*. In other words, if* $j_r = j_{r+1} = j_{r+2}$ *it is not possible to have* $j_{r+3} = j_{r+4}$.

Proof. From Corollary 3, we know that

if $j_r = j_{r+1} = j_{r+2}$ then $S_{r+1}[j_{r+1}] = S_{r+2}[i_{r+2}] = S_{r+2}[j_{r+2}] = 0$

So $S_{r+2}[i_{r+2} + 1]$ cannot be 0 and hence j_{r+3} must be different from j_{r+2}.

Therefore, $S_{r+3}[i_{r+3}]$ as well as $S_{r+3}[j_{r+3}]$ would be non-zero (please note that it is possible for both i and j to be same in this round).

Next, $(i_{r+3} + 1)$ cannot be same as i_{r+2} for $N > 2$. This implies, $S_{r+3}[i_{r+3} + 1]$ cannot be 0.

Thus, j_{r+4} cannot be equal to j_{r+3}. Hence the proof. □

Theorem 4. *During RC4 PRGA, there cannot be a continuously decreasing sequence of* j *of length more than* 3 *or in other words there cannot exist any* r *for which* $(j_r - j_{r+1}) = (j_{r+1} - j_{r+2}) = (j_{r+2} - j_{r+3}) = k$ *where* $(k < N - 1)$.

Proof. Let us consider an r such that -

$(j_r - j_{r+1}) = (j_{r+1} - j_{r+2}) = k$ where $(k < N - 1)$

Since, $(j_r - j_{r+1}) = k$, it can be said that $S_r[i_r + 1] = (N - k)$.

Similarly, as $(j_{r+1} - j_{r+2}) = k$, $S_{r+1}[i_{r+1} + 1] = (N - k)$.

From the above two equations, it can be concluded that - $i_{r+1} = j_{r+1} - 1$

In RC4 PRGA, since a Finney cycle cannot happen, k cannot take the value of $(N-1)$ and hence $(k < N-1)$.

After the swap operation is completed in round $(r+2)$, $S_{r+2}[j_{r+2}] = (N-k)$.

Also, we know that $(j_{r+1} - j_{r+2}) = k$ which implies $j_{r+2} = j_{r+1} + (N-k)$.

As $(k < N-1)$, it is not possible to have $j_{r+2} = j_{r+1} + 1$.

So, i_{r+2} cannot be one less than j_{r+2} implying it would not be possible to have $(j_{r+2} - j_{r+3}) = k$. Hence the proof. □

Corollary 6. *During RC4 PRGA, there cannot be a continuously increasing sequence of j of length more than 3 or in other words there cannot exist any r for which $(j_{r+1} - j_r) = (j_{r+2} - j_{r+1}) = (j_{r+3} - j_{r+2}) = k$ where $(k > 1)$.*

Proof. Any increase of a j value between two successive rounds by k can be considered as a decrease of the j value between two successive rounds by $(N-k)$. Here, $(k > 1)$. Hence, we may apply the previous theorem and arrive at this result. □

To illustrate Theorem 4 and Corollary 6, once can say that the following sequences of j can never happen in RC4 cycle (where $N = 8$).

For example, increasing sequences like $(1,3,5,7)$ or $(1,4,7,2)$ or $(1,5,1,5)$ etc. and decreasing sequences like $(7,6,5,4)$ or $(7,5,3,1)$ or $(7,4,1,6)$ etc. can never happen in RC4.

Theorem 5. *In RC4, there must be more than one non-Finney cycles.*

Proof. We prove the above claim by contradiction.

Let us assume that for some value of N (of the form 2^n) there exists only 1 non-Finney cycle.

Total number of permutations (of S-Box of length N together with i and j) is $(N!).(N^2)$.

Hence, by eliminating the permutations pertaining to Finney cycle, we get that the number of permutations as part of the non-Finney cycle would be

$T = (N!).(N^2) - (N!)$

We now use the results derived by Mister & Tavares([6]).

Let us denote by S_0, any permutation of S-Box where $i = 0$ and $j = 0$. Since there is only 1 non-Finney cycle, all possible $(N-1)$ right shifts of S_0 must be part of the same cycle. Let the right-shifts appear in the cycle as per the sequence $d_1, d_2, ..., d_{N-1}$ where $d_0 (= d_N)$ is understood to be same 0 (no-shift). As per the Cycle Partitioning Theorem ([6]) one can say that if L is the distance (number of rounds required to reach d_{i+1} from d_i), then

$L = \frac{T}{N} = (N-1)!.(N^2) - (N-1)!$

Let us denote d_1 by k (where $1 <= k <= (N-1)$). Also, let S_1 be the permutation of S-box corresponding to the k-shift.

S_0 starts with $i = 0$, hence after k shifts $i = k$. So $(L-k)$ must be divisible by N - since k steps before S_1, i must be equal to 0.

Now $(L-k)$ is same as $[(T/N) - k]$ which equals $(N-1)!.(N^2) - (N-1)! - k$

Since N is of the form 2^n, the first two terms of the above expression would be divisible by N. But the third term would not be divisible by N.

Hence we reach a contradiction. This implies our initial assumption is incorrect.

Therefore, in RC4, there must be more than 1 non-Finney cycles. □

We now prove the following theorem by extending the Cycle Partitioning results of Mister & Tavares([6]).

Theorem 6. *Let S_0 be the initial permutation of S-Box ($i = 0$, $j = 0$) in an RC4 cycle with right-shifts that appear as per sequence of $(d_1, d_2, ..., d_{k-1})$ where $d_0 (= d_k)$ represents the original permutation or in other words a complete rotation by N-bytes. In that case,*

1. *$k = \frac{LCM(N, d_1)}{d_1}$*
2. *k must be of the form 2^m where possible values for m could be $(1, 2, ..., N)$*
3. *If T is the length of the cycle that starts with $(S_0, i = 0, j = 0)$, and $(m > 0)$, then there must be at least $\frac{N}{2^m} = 2^(n-m)$ disjoint cycles in the state space of cycle length T*
4. *$(\frac{T}{k} - d_1)$ must be divisible by N apart from the condition that T must be divisible by N*

Proof. The claim for each part of the theorem is proved below -

1. As d_1 is the first shifted occurrence in the cycle, it is clear that after repeating the same same shift k number of times it must come back to the original permutation. Hence, k times d_1 must be a multiple of N. Again the multiple must be the least one as otherwise the permutation would have come back to the original state even before the length of T which is not possible. Therefore, the product of k and d_1 must be same as $LCM(N, d_1)$. Hence, the result.

2. Since d_1 times k is of the form $LCM(N, d_1)$ and N is of the form 2^n, k must be of the form 2^m where possible values for m could be $(1, 2, ..., N)$

3. Since k is of the form 2^m where $(m > 0)$, there would be $2^(n - m)$ disjoint cycles in the state space of cycle length T. Please note that it is possible that some of these cycles may never appear in real RC4 cycles as they may not contain any permutation corresponding to $(i = 0, j = 0)$. The reason of qualifying the result with "at least" is to highlight that there may be other cycles in RC4 state space that somehow have the same length as the current cycle under consideration.
4. When the permutation reaches the first right shift of d_1, value of index i would also be d_1. At this point, the number of rounds that would have occurred is $\frac{T}{k}$ and d_1 rounds earlier to this point the value of i must have been 0.
   Hence, $(\frac{T}{k} - d_1)$ must be divisible by N apart from the condition that T must be divisible by N

If we assume that the RC4 state traversal is equivalent to a random traversal (although this is not strictly true - since not all states can be reached in real RC4 starting with the condition of $i = 0$ and $j = 0$), the chance of S_0 arriving first at right shifted permutation of 1 byte in RC4 cycle, is only 1 out of N. This implies, the state space of RC4 is more likely to consist of multiple cycles of same length than a few cycles of unique lengths.

4 Conclusion

Rewrite

The pseudo-randomness of the index j in RC4 has been an open question for quite some time. In this note we show that

j is indeed not pseudo-random in long term evolution of RC4 PRGA where we consider S as a pseudo-random permutation.

To the best of our knowledge, this result has not been noted earlier. The implication of this result could be interesting

to obtain further non-randomness in the evolution of RC4. Moreover, the result may be utilized to obtain additional biases

at the initial stage of RC4 PRGA where the permutation S has certain non-randomness.

References

1. H. Finney.
   An RC4 cycle that can't happen. Post in sci.crypt, September 1994.
2. R. J. Jenkins. ISAAC and RC4. 1996.
   Available at `http://burtleburtle.net/bob/rand/isaac.html`
   [last accessed on October 25, 2015].

3. K. G. Paterson, B. Poettering and J. C. N. Schuldt.
   Big Bias Hunting in Amazonia: Large-scale Computation and Exploitation of RC4 Biases. ASIACRYPT 2014.
   LNCS, Part 1, pp. 398–419, Vol. 8873, 2014.
4. S. SenGupta, S. Maitra, G. Paul, S. Sarkar.
   (Non–)Random Sequences from (Non–)Random Permutations – Analysis of RC4 stream cipher.
   Journal of Cryptology, 27(1):67–108, 2014
5. P. Sepehrdad, S. Vaudenay, and M. Vuagnoux.
   Statistical Attack on RC4 - Distinguishing WPA.
   EUROCRYPT 2011. LNCS pp. 343–363, Vol. 6632, 2011.
6. S. Mister and S. E. Tavares.
   Cryptanalysis of RC4-like Ciphers. International Workshop on Selected Areas in Cryptography.
   SAC 1998, pp. 131-143, 1998.