# The Index $j$ in RC4 is not Pseudo-random due to Non-existence of Finney Cycle

Subhamoy Maitra
Applied Statistics Unit, Indian Statistical Institute, Kolkata, India
Email: subho@isical.ac.in

**Abstract**

In this very short note we prove that the pseudo-random index $j$ of RC4 is indeed not pseudo-random. This is a simple result that missed our attention for quite a long time. 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.

**Keywords:** RC4, Non-randomness. Pseudo-random Index.

## 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, 4, 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 $N - 1$:

$$j = j + S[i] + K[i \bmod l];\ \mathrm{swap}(S[i], S[j]);$$

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

- $i = j = 0$;
- for $i = 0$ to $N - 1$:

$$i = i + 1;\ j = j + S[i];\ \mathrm{swap}(S[i], S[j]);\ z = S[S[i] + S[j]];$$

Note that all the additions here are modulo $N$.

## 2 Proof of the result

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

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

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:

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

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 Conclusion

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.