Title	New Integrated Long-Term Glimpse of RC4
Author(s)	Ito, Ryoma; Miyaji, Atsuko
Citation	Lecture Notes in Computer Science: 137-149
Issue Date	2015-01-22
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/13006
Rights	This is the author-created version of Springer, Ryoma Ito and Atsuko Miyaji, Lecture Notes in Computer Science, 2015, pp.137-149. The original publication is available at www.springerlink.com, http://dx.doi.org/10.1007/978-3-319-15087-1_11
Description	15th International Workshop, WISA 2014, Jeju Island, Korea, August 25-27, 2014. Revised Selected Papers

New Integrated Long-Term Glimpse of RC4

Ryoma Ito and Atsuko Miyaji

Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi-shi, Ishikawa, 923-1292, Japan
{s1310005,miyaji}@jaist.ac.jp

Abstract. RC4, which was designed by Ron Rivest in 1987, is widely used in various applications such as SSL/TLS, WEP, WPA, etc. In 1996, Jenkins discovered correlations between one output keystream and a state location, known as Glimpse Theorem. In 2013, Maitra and Sen Gupta proved Glimpse Theorem and showed correlations between two consecutive output keystreams and a state location, called long-term Glimpse. In this paper, we show a new long-term Glimpse and integrate both the new and the previous long-term Glimpse into a whole.

Keywords: RC4, correlation, long-term Glimpse

1 Introduction

RC4, which was designed by Ron Rivest in 1987, is widely used in various applications such as Secure Socket Layer/Transport Layer Security (SSL/TLS), Wired Equivalent Privacy (WEP) and Wi-fi Protected Access (WPA), etc. Due to its popularity and simplicity, RC4 has become a hot cryptanalysis target since its specification was made public on the internet in 1994. For example, typical attacks on RC4 are distinguishing attack [3, 4, 10], state recovery attack [1, 6, 9] and key recovery attack [2, 8, 11].

In 1996, Jenkins discovered correlations between one output keystream and a state location, which is known as Glimpse Theorem [5]. These correlations have biases with the probability about $\frac{2}{N}$ higher than that of random association $\frac{1}{N}$ using the knowledge of one output keystream. In 2013, Maitra and Sen Gupta presented the complete proof of Glimpse Theorem and showed $S_r[r+1] = N-1$ occurs with the probability about $\frac{2}{N}$ when two consecutive output keystreams Z_r and Z_{r+1} satisfies $Z_{r+1} = Z_r$, where $S_r[r+1]$ is the $r+1$-th location of the state array in the r-th round as usual. They also showed the probability of $S_r[r+1] = N-1$ is further increased to about $\frac{3}{N}$ when $Z_{r+1} = r+2$ as well as $Z_{r+1} = Z_r$ occurs. Here, we call correlation with a probability significantly higher or lower than $\frac{1}{N}$ (the probability of random association) *positive bias* or *negative bias*, respectively. Then, their results of $S_r[r+1] = N-1$ with the probability about $\frac{2}{N}$ correspond to cases with positive biases. Note that Theorem 2 implicitly means that there exists a value of $S_r[r+1]$ with negative bias since $S_r[r+1]$ varies in $[0, N-1]$ when $Z_{r+1} = Z_r$ has happened. We often assume uniform randomness of other certain events to prove bias of a certain event.

Therefore, it is important to prove the existence of a value with negative bias explicitly. We also call such a case with negative bias to *dual case* of a positive bias.

In this paper, we first show a dual case of $S_r[r+1] = N-1$, that is $S_r[r+1] = 0$, occurs with the probability about $\frac{1}{N^2}$ when $Z_{r+1} = Z_r$, which will be shown as Theorem 4. Then, Theorem 5 will give each probability of $S_r[r+1] = 0$ when $Z_{r+1} = r + x$ ($\forall x \in [0, N-1]$) as well as $Z_{r+1} = Z_r$ occurs. Furthermore, during our careful observation of the dual case, we also find a new positive bias on $S_r[r+1]$, which will be shown in Theorem 6. Our results show that, giving two consecutive keystreams Z_r and Z_{r+1} satisfying with $Z_{r+1} = Z_r$ and $Z_{r+1} = r + 1 + x$ ($x \in [2, N-1]$), the probability of $S_r[r+1] = N - x$ is about $\frac{2}{N}$, which is significantly higher than random association $\frac{1}{N}$. Note that the previous results are limited to a value of $S_r[r+1] = N-1$, but our results varies $S_r[r+1] \in [0, N-2]$. Furthermore, both our new and the previous results are integrated into long-term Glimpse of $Z_{r+1} = Z_r$ in Theorem 7.

This paper is organized as follows. Section 2 briefly summarizes notation and RC4 algorithms. Section 3 presents the previous works on Glimpse Theorem [5] and long-term Glimpse [7]. Section 4 first discusses positive and negative biases, and shows Theorems 4 to 7. Section 5 demonstrates experimental simulations. Section 6 concludes this paper.

2 Preliminary

The following notation is used in this paper.

K, l : secret key, the length of secret key (bytes)
r : number of rounds
N : number of arrays in state (typically $N = 256$)
S_r^K or S_r : state of KSA or PRGA after the swap in the r-th round
i_r, j_r : indices of S_r for the r-th round
Z_r : one output keystream for the r-th round
t_r : index of Z_r

RC4 consists of two algorithms: Key Scheduling Algorithm (KSA) and Pseudo Random Generation Algorithm (PRGA). KSA generates the state S_N^K from a secret key K of l bytes as described in Algorithm 1. Then, the final state S_N^K in KSA becomes the input of PRGA as S_0. Once the state S_0 is computed, PRGA generates one output keystream Z_r of bytes as described in Algorithm 2. The output keystream Z_r will be XORed with a plaintext to generate a ciphertext.

Algorithm 1 KSA

for $i = 0$ to $N-1$ **do**
  $S_0^K[i] \leftarrow i$
end for
$j \leftarrow 0$
for $i = 0$ to $N-1$ **do**
  $j \leftarrow j + S_i^K[i] + K[i \mod l]$
  Swap$(S_i^K[i], S_i^K[j])$
end for

Algorithm 2 PRGA

$r \leftarrow 0$, $i_0 \leftarrow 0$, $j_0 \leftarrow 0$
loop
  $r \leftarrow r+1$, $i_r \leftarrow i_{r-1} + 1$
  $j_r \leftarrow j_{r-1} + S_{r-1}[i_r]$
  Swap$(S_{r-1}[i_r], S_{r-1}[j_r])$
  $t_r \leftarrow S_r[i_r] + S_r[j_r]$
  Output: $Z_r \leftarrow S_r[t_r]$
end loop

In this paper, we focus on PRGA and investigate correlations between two consecutive output keystreams and a state location. The probability of one location by random association is $\frac{1}{N}$ and uniform randomness of the RC4 stream cipher is assumed if there are no significant biases.

3 Previous works

In 1996, Jenkins discovered correlations between one output keystream and a state location [5], which is proved as Glimpse Theorem in [7]. Glimpse Theorem is given as follows.

Theorem 1. **[7]** *After the r-th round of PRGA for $r \geq 1$, we have*

$$\Pr(S_r[j_r] = i_r - Z_r) = \Pr(S_r[i_r] = j_r - Z_r) \approx \frac{2}{N}.$$

In 2013, Maitra and Sen Gupta discovered other correlations between two consecutive output keystreams and the $r+1$-th location of the state array in the r-th round, which is called long-term Glimpse [7]. Long-term Glimpse is given as follows. Note that Theorem 3 is a special case of Theorem 2.

Theorem 2. **[7]** *After the r-th round of PRGA for $r \geq 1$, we have*

$$\Pr(S_r[r+1] = N-1 | Z_{r+1} = Z_r) \approx \frac{2}{N}.$$

Theorem 3. **[7]** *After the r-th round of PRGA for $r \geq 1$, we have*

$$\Pr(S_r[r+1] = N-1 | Z_{r+1} = Z_r \wedge Z_{r+1} = r+2) \approx \frac{3}{N}.$$

4 New results on long-term Glimpse

4.1 Observation

Let us investigate the previous results (Theorems 2 and 3) in detail. Here, we call correlation with a probability significantly higher or lower than $\frac{1}{N}$ (the probability of random association) to *positive bias* or *negative bias*, respectively. Theorems 2 and 3 give cases with positive biases. Then, Theorem 2 implicitly means that there exists a value of $S_r[r+1]$ with negative bias since $S_r[r+1]$ varies in $[0, N-1]$ even when $Z_{r+1} = Z_r$ has happened. We often assume uniform randomness of other certain events to prove bias of a certain event. Therefore, it is important to prove the existence of a value in $S_r[r+1]$ with negative bias explicitly. We also call such a case with negative bias a *dual case* of a positive bias.

One of our motivation is to find a dual case of Theorem 2, which will be shown as Theorem 4. Then, we will also prove a special case of Theorem 4 in the same way as Theorem 3 to Theorem 2, which will be shown as Theorem 5. Furthermore, during our careful observation of the dual case, we also find a new positive bias on $S_r[r+1]$, which will be shown in Theorem 6. Our new results can integrate long-term Glimpse when $Z_{r+1} = Z_r$. The previous results are limited to the case of $S_r[r+1] = N-1$ when $Z_{r+1} = Z_r$. Our results are not limited to $S_r[r+1] = N-1$ but varies $S_r[r+1] \in [0, N-2]$. Finally, both results can be integrated in Theorem 7.

4.2 New negative biases

First, Theorem 4 shows a dual case of Theorem 2 as follows.

Theorem 4. *After the r-th round of PRGA for $r \geq 1$, we have*

$$\Pr(S_r[r+1] = 0 | Z_{r+1} = Z_r) \approx \frac{2}{N^2}\left(1 - \frac{1}{N}\right).$$

Proof. We define main events as follows:

$$A := (S_r[r+1] = 0), B := (Z_{r+1} = Z_r).$$

We first compute $\Pr(B|A)$, and apply Bayes' theorem to prove the claim. Assuming that event A happened, we get

$$j_{r+1} = j_r + S_r[i_{r+1}] = j_r + S_r[r+1] = j_r.$$

Then, $\Pr(B|A)$ is computed in three paths: $j_r = r$ (Path 1), $j_r = r+1$ (Path 2) and $j_r \neq r, r+1$ (Path 3). These paths include all events in order to compute $\Pr(B|A)$. Let $X = S_r[r]$ and $Y = S_r[j_r]$.

Path 1. Fig. 1 shows a state transition diagram in Path 1. First, we prove $t_r \neq t_{r+1}$. After the r-th round, $t_r = 2X$ holds since $i_r = j_r = r$. In the next round, $t_{r+1} = X$ holds since $j_{r+1} = j_r = r$ and $i_{r+1} = r+1$. Thus, we get $t_r \neq t_{r+1}$ with probability 1 since $X \neq 0$. Then, if event B occurs, t_{r+1} must be swapped from t_r. This is why $\Pr(\text{Path 1}) = \Pr(B|A \wedge j_r = r)$ is computed in two subpaths: $i_r = 1 \wedge t_{r+1} = 1$ (Path 1-1) and $i_r = 254 \wedge t_{r+1} = 255$ (Path 1-2).

Path 1-1. Fig. 2 shows a state transition diagram in Path 1-1. Then, we get event B since $Z_{r+1} = S_{r+1}[1] = 0$ and $Z_r = S_r[2] = 0$. Thus, we can compute the probability of Path 1-1 as follows.

$$\Pr(\text{Path 1-1}) = \Pr(\text{Path 1} \wedge i_r = 1 \wedge t_{r+1} = 1) = 1.$$

Path 1-2. Fig. 3 shows a state transition diagram in Path 1-2. Then, we get event B since $Z_{r+1} = S_{r+1}[255] = 255$ and $Z_r = S_r[254] = 255$. Thus, we can compute the probability of Path 1-2 as follows.

$$\Pr(\text{Path 1-2}) = \Pr(\text{Path 1} \wedge i_r = 254 \wedge t_{r+1} = 255) = 1.$$

Therefore, the probability of Path 1 is computed as follows.

$$\begin{aligned}\Pr(\text{Path 1}) = {} & \Pr(\text{Path 1-1}) \cdot \Pr(i_r = 1 \wedge t_{r+1} = 1) \\ & + \Pr(\text{Path 1-2}) \cdot \Pr(i_r = 254 \wedge t_{r+1} = 255) \\ \approx {} & 1 \cdot \left(\frac{1}{N} \cdot \frac{1}{N}\right) + 1 \cdot \left(\frac{1}{N} \cdot \frac{1}{N}\right) = \frac{2}{N^2}.\end{aligned}$$

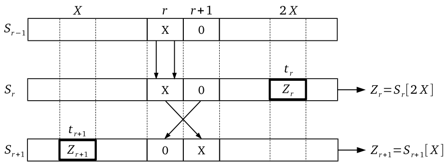

Fig. 1. Path 1

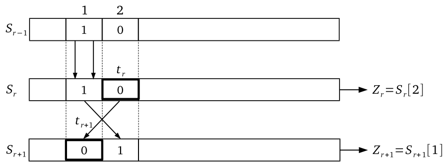

Fig. 2. Path 1-1

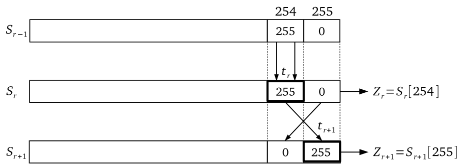

Fig. 3. Path 1-2

Path 2. Fig. 4 shows a state transition diagram in Path 2. We get $t_r \neq t_{r+1}$ in the same way as Path 1. Then, event B never occurs because t_{r+1} can not be swapped from t_r. Therefore, the probability of Path 2 is computed as follows.

$$\Pr(\text{Path 2}) = \Pr(B|A \wedge j_r = r+1) = 0.$$

Path 3. Fig. 5 shows a state transition diagram in Path 3. We get $t_r \neq t_{r+1}$ in the same way as Path 1. Then, if event B occurs, t_{r+1} must be swapped from t_r. This is why $\Pr(\text{Path 3}) = \Pr(B|A \wedge j_r \neq r, r+1)$ is computed in two subpaths: $t_r = j_r \wedge t_{r+1} = r+1$ (Path 3-1) and $t_r = r+1 \wedge t_{r+1} = j_{r+1}$ (Path 3-2).

Path 3-1. Fig. 6 shows a state transition diagram in Path 3-1. Then, we get event B since $Z_{r+1} = S_{r+1}[r+1] = r+1$ and $Z_r = S_r[j_r] = r+1$. Thus, we can compute the probability of Path 3-1 as follows.

$$\Pr(\text{Path 3-1}) = \Pr(\text{Path 3} \wedge t_r = j_r \wedge t_{r+1} = r+1) = 1.$$

Path 3-2. Fig. 7 shows a state transition diagram in Path 3-2. Then, we get event B since $Z_{r+1} = S_{r+1}[j_{r+1}] = 0$ and $Z_r = S_r[r+1] = 0$. Thus, we can compute the probability of Path 3-2 as follows.

$$\Pr(\text{Path 3-2}) = \Pr(\text{Path 3} \wedge t_r = r+1 \wedge t_{r+1} = j_r) = 1.$$

Therefore, the probability of Path 3 is computed as follows.

$$\begin{aligned}\Pr(\text{Path 3}) &= \Pr(\text{Path 3-1}) \cdot \Pr(t_r = j_r \wedge t_{r+1} = r+1) \\ &\quad + \Pr(\text{Path 3-2}) \cdot \Pr(t_r = r+1 \wedge t_{r+1} = j_{r+1}) \\ &\approx 1 \cdot \left(\frac{1}{N} \cdot \frac{1}{N}\right) + 1 \cdot \left(\frac{1}{N} \cdot \frac{1}{N}\right) = \frac{2}{N^2}.\end{aligned}$$

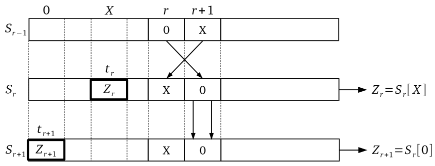

Fig. 4. Path 2

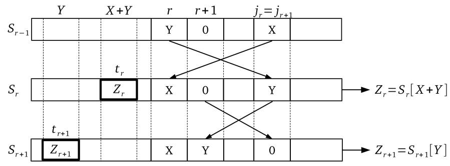

Fig. 5. Path 3

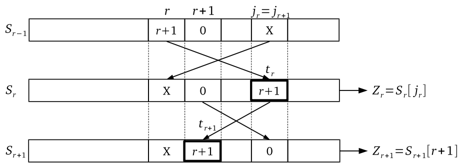

Fig. 6. Path 3-1

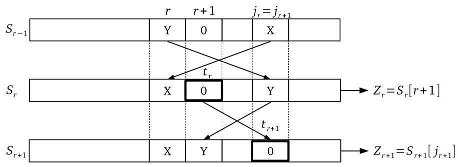

Fig. 7. Path 3-2

From these results, $\Pr(B|A)$ is computed as follows.

$$\begin{aligned}\Pr(B|A) &= \Pr(\text{Path 1}) \cdot \Pr(j_r = r) + \Pr(\text{Path 2}) \cdot \Pr(j_r = r+1) \\ &\quad + \Pr(\text{Path 3}) \cdot \Pr(j_r \neq r, r+1) \\ &\approx \frac{2}{N^2} \cdot \frac{1}{N} + 0 \cdot \frac{1}{N} + \frac{2}{N^2} \cdot \left(1 - \frac{2}{N}\right) = \frac{2}{N^2}\left(1 - \frac{1}{N}\right).\end{aligned}$$

$\Pr(A|B)$ is computed as follows by applying Bayes' theorem since events A and B occur with the probability of random association $\frac{1}{N}$.

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)} \approx \frac{\frac{2}{N^2}\left(1 - \frac{1}{N}\right) \cdot \frac{1}{N}}{\frac{1}{N}} = \frac{2}{N^2}\left(1 - \frac{1}{N}\right). \qquad \square$$

Next, Theorem 5 shows a special case of Theorem 4 as follows.

Theorem 5. *After the r-th round of PRGA for $r \geq 1$ and $\forall x \in [0, N-1]$, we have*

$$\Pr(S_r[r+1] = 0 | Z_{r+1} = Z_r \wedge Z_{r+1} = r + x) \approx \begin{cases} \frac{1}{N}\left(1 - \frac{2}{N^2}\right) & \textit{if } x = 1 \\ \frac{2}{N^2}\left(1 - \frac{1}{N}\right) & \textit{if } x = 255 \\ \frac{1}{N^2}\left(1 - \frac{2}{N}\right) & \textit{if } x = N - r \\ & (x \neq 1, 255). \end{cases}$$

Proof. We define main events as follows.

$$A := (S_r[r+1] = 0), B := (Z_{r+1} = Z_r), C := (Z_{r+1} = r + x).$$

$\Pr(A|B \wedge C)$ is difficult to compute because events B and C are not independent. To avoid this problem, we define a new event $B' := (Z_r = r+x)$. Then, $\Pr(A|B \wedge C) = \Pr(A|B' \wedge C)$ since $B \wedge C$ and $B' \wedge C$ are the same event. $\Pr(A|B' \wedge C)$ is decomposed as follows by using Bayes' theorem:

$$\Pr(A|B' \wedge C) = \frac{\Pr(A \wedge B' \wedge C)}{\Pr(B' \wedge C)} = \frac{\Pr(C|B' \wedge A) \cdot \Pr(B'|A) \cdot \Pr(A)}{\Pr(B' \wedge C)}.$$

We first compute $\Pr(C|B' \wedge A)$ in three paths: $j_r = r$ (Path 1), $j_r = r+1$ (Path 2) and $j_r \neq r, r+1$ (Path 3). These paths are the same as in Theorem 4, and thus the proof itself is similar to Theorem 4. Let $X = S_r[r]$ and $Y = S_r[j_r]$.

Path 1. Fig. 1 shows a state transition diagram in Path 1. Note that $t_r \neq t_{r+1}$ from the discussion of Path 1 in Theorem 4, and that event C is limited to two subpaths: $i_r = 1$ for $r+x = 0$ (Path 1-1) and $t_{r+1} = 255$ for $r+x = 255$ (Path 1-2).

Path 1-1. Fig. 2 shows a state transition diagram in Path 1-1. Then, event C holds under event $B' \wedge A$ since $Z_{r+1} = S_{r+1}[1] = 0$ and $Z_r = S_r[2] = 0$. Note that $i_r = 1$ and $r+x = 0$ hold if and only if $x = 255$. Thus, we can compute the probability of Path 1-1 as follows.

$$\Pr(\text{Path 1-1}) = \Pr(\text{Path 1} \wedge i_r = 1) = 1 \quad \text{if } x = 255.$$

Path 1-2. Fig. 3 shows a state transition diagram in Path 1-2. Then, event C holds under event $B' \wedge A$ since $Z_{r+1} = S_{r+1}[255] = 255$ and $Z_r = S_r[254] = 255$. Note that $i_r = 254$ (see Fig. 3) and $r+x = 255$ hold if and only if $x = 1$. Thus, we can compute the probability of Path 1-2 as follows.

$$\Pr(\text{Path 1-2}) = \Pr(\text{Path 1} \wedge t_{r+1} = 255) = 1 \quad \text{if } x = 1.$$

Therefore, the probability of Path 1 is computed as follows.

$$\Pr(\text{Path 1}) = \begin{cases} \Pr(\text{Path 1-1}) \cdot \Pr(i_r = 1) \approx \dfrac{1}{N} & \text{if } x = 255 \\ \Pr(\text{Path 1-2}) \cdot \Pr(t_{r+1} = 255) \approx \dfrac{1}{N} & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Path 2. Event C never occurs in Path 2 from the discussion of Path 2 in Theorem 4. Therefore, the probability of Path 2 is computed as follows.

$$\Pr(\text{Path 2}) = \Pr(C|B' \wedge A \wedge j_r = r+1) = 0.$$

Path 3. Fig. 5 shows a state transition diagram in Path 3. Note that $t_r \neq t_{r+1}$ from the discussion of Path 3 in Theorem 4, and that event C is limited to two subpaths: $t_{r+1} = r+1$ for $x = 1$ (Path 3-1) and $t_r = r+1 \wedge t_{r+1} = j_{r+1}$ for $r+x = 0$ (Path 3-2).

Path 3-1. Fig. 6 shows a state transition diagram in Path 3-1. Then, event C holds under event $B' \wedge A$ since $Z_{r+1} = S_{r+1}[r+1] = r+1$ and $Z_r = S_r[j_r] = r+1$. Thus, we can compute the probability of Path 3-1 as follows.

$$\Pr(\text{Path 3-1}) = \Pr(\text{Path } 3 \wedge t_{r+1} = r+1) = 1 \quad \text{if } x = 1.$$

Path 3-2. Fig. 7 shows a state transition diagram in Path 3-2. Then, event C holds under event $B' \wedge A$ since $Z_{r+1} = S_{r+1}[j_{r+1}] = 0$ and $Z_r = S_r[r+1] = 0$. Note that $r + x = 0$ ($\forall r \in [0, N-1]$) means $x = N - r$. Thus, we can compute the probability of Path 3-2 as follows.

$$\Pr(\text{Path 3-2}) = \Pr(\text{Path } 3 \wedge t_r = r+1 \wedge t_{r+1} = j_{r+1}) = 1.$$

Therefore, the probability of Path 3 is computed as follows.

$$\begin{aligned}\Pr(\text{Path 3}) &= \Pr(\text{Path 3-1}) \cdot \Pr(t_{r+1} = r+1) \\ &\quad + \Pr(\text{Path 3-2}) \cdot \Pr(t_r = r+1 \wedge t_{r+1} = j_{r+1}) \\ &\approx \begin{cases} 1 \cdot \dfrac{1}{N} + 1 \cdot \left(\dfrac{1}{N} \cdot \dfrac{1}{N}\right) = \dfrac{1}{N}\left(1 + \dfrac{1}{N}\right) & \text{if } x = 1 \\ 0 \cdot \dfrac{1}{N} + 1 \cdot \left(\dfrac{1}{N} \cdot \dfrac{1}{N}\right) = \dfrac{1}{N^2} & \text{if } x = N - r \ (x \neq 1). \end{cases}\end{aligned}$$

From these results, $\Pr(C|B' \wedge A)$ is computed as follows.

$$\begin{aligned}\Pr(C|B' \wedge A) &= \Pr(\text{Path 1}) \cdot \Pr(j_r = r) + \Pr(\text{Path 2}) \cdot \Pr(j_r = r+1) \\ &\quad + \Pr(\text{Path 3}) \cdot \Pr(j_r \neq r, r+1) \\ &\approx \begin{cases} \dfrac{1}{N} \cdot \dfrac{1}{N} + \dfrac{1}{N} \cdot \left(1 + \dfrac{1}{N}\right) \cdot \left(1 - \dfrac{2}{N}\right) = \dfrac{1}{N}\left(1 - \dfrac{2}{N^2}\right) & \text{if } x = 1 \\ \dfrac{1}{N} \cdot \dfrac{1}{N} + \dfrac{1}{N^2} \cdot \left(1 - \dfrac{2}{N}\right) = \dfrac{2}{N^2}\left(1 - \dfrac{1}{N}\right) & \text{if } x = 255 \\ 0 \cdot \dfrac{1}{N} + \dfrac{1}{N^2} \cdot \left(1 - \dfrac{2}{N}\right) = \dfrac{1}{N^2}\left(1 - \dfrac{2}{N}\right) & \text{if } x = N - r \\ & (x \neq 1, 255). \end{cases}\end{aligned}$$

$\Pr(A|B \wedge C)$ is computed as follows by applying Bayes' theorem since events A, B', C and $B'|A$ occur with the probability of random association $\frac{1}{N}$.

$$\begin{aligned}\Pr(A|B \wedge C) &= \frac{\Pr(C|B' \wedge A) \cdot \Pr(B'|A) \cdot \Pr(A)}{\Pr(B' \wedge C)} \approx \frac{\Pr(C|B' \wedge A) \cdot \frac{1}{N} \cdot \frac{1}{N}}{\frac{1}{N} \cdot \frac{1}{N}} \\ &= \Pr(C|B' \wedge A) \approx \begin{cases} \dfrac{1}{N}\left(1 - \dfrac{2}{N^2}\right) & \text{if } x = 1 \\ \dfrac{2}{N^2}\left(1 - \dfrac{1}{N}\right) & \text{if } x = 255 \\ \dfrac{1}{N^2}\left(1 - \dfrac{2}{N}\right) & \text{if } x = N - r \ (x \neq 1, 255). \end{cases} \qquad \square\end{aligned}$$

4.3 New positive biases and their integration

Theorem 6 shows a new positive bias on $S_r[r+1]$ as follows.

Theorem 6. *After the r-th round of PRGA for $r \geq 1$ and $\forall x \in [2, N-1]$, we have*

$$\Pr(S_r[r+1] = N - x | Z_{r+1} = Z_r \wedge Z_{r+1} = r+1+x) \approx \frac{2}{N}\left(1 - \frac{1}{N} + \frac{1}{N^2}\right).$$

Proof. We define main events as follows.

$$A := (S_r[r+1] = N - x), B := (Z_{r+1} = Z_r),$$
$$B' := (Z_r = r+1+x), C := (Z_{r+1} = r+1+x).$$

The proof itself is similar to Theorem 5. We first compute $\Pr(C|B' \wedge A)$ in three paths: $j_r = r$ (Path 1), $j_r = r+1$ (Path 2) and $j_r \neq r, r+1$ (Path 3). Let $X = S_r[r]$, $Y = S_r[j_r]$ and $W = S_r[j_{r+1}]$.

Path 1. Both t_r and t_{r+1} are independent since we get $t_r = 2X$ and $t_{r+1} = N - x + W$. Then, event C is limited to three subpaths: $t_{r+1} = r+1$ (Path 1-1), $N - x = r+1+x \wedge t_{r+1} = j_{r+1}$ (Path 1-2) and $t_{r+1} = t_r$ except when t_r equals either $r+1$ or j_{r+1} (Path 1-3). We can compute the probability of each subpath as follows.

$$\Pr(\text{Path 1-1}) = \Pr(\text{Path 1} \wedge t_{r+1} = r+1) = 1,$$
$$\Pr(\text{Path 1-2}) = \Pr(\text{Path 1} \wedge N - x = r+1+x \wedge t_{r+1} = j_{r+1}) = 1,$$
$$\Pr(\text{Path 1-3}) = \Pr(\text{Path 1} \wedge t_{r+1} = t_r) = 1 - \tfrac{2}{N}.$$

Therefore, the probability of Path 1 is computed as follows.

$$\begin{aligned}
\Pr(\text{Path 1}) = \Pr(\text{Path 1-1}) \cdot \Pr(t_{r+1} = r+1) \\
+ \Pr(\text{Path 1-2}) \cdot \Pr(N - x = r+1+x \wedge t_{r+1} = j_{r+1}) \\
+ \Pr(\text{Path 1-3}) \cdot \Pr(t_{r+1} = t_r) \\
\approx 1 \cdot \frac{1}{N} + 1 \cdot \left(\frac{1}{N} \cdot \frac{1}{N}\right) + \left(1 - \frac{2}{N}\right) \cdot \frac{1}{N} = \frac{1}{N}\left(2 - \frac{1}{N}\right).
\end{aligned}$$

Path 2. We get $t_r \neq t_{r+1}$ since $t_r = N - x + X$, $t_{r+1} = N - x + W$ and $X \neq W$. Then, event C is limited to two subpaths: $t_{r+1} = r+1$ (Path 2-1) and $N - x = r+1+x \wedge t_{r+1} = j_{r+1}$ (Path 2-2). We can compute the probability of each subpath as follows.

$$\Pr(\text{Path 2-1}) = \Pr(\text{Path 2} \wedge t_{r+1} = r+1) = 1,$$
$$\Pr(\text{Path 2-2}) = \Pr(\text{Path 2} \wedge N - x = r+1+x \wedge t_{r+1} = j_{r+1}) = 1.$$

Therefore, the probability of Path 2 is computed as follows.

$$\begin{aligned}
\Pr(\text{Path 2}) = \Pr(\text{Path 2-1}) \cdot \Pr(t_{r+1} = r+1) \\
+ \Pr(\text{Path 2-2}) \cdot \Pr(N - x = r+1+x \wedge t_{r+1} = j_{r+1}) \\
\approx 1 \cdot \frac{1}{N} + 1 \cdot \left(\frac{1}{N} \cdot \frac{1}{N}\right) = \frac{1}{N}\left(1 + \frac{1}{N}\right).
\end{aligned}$$

Path 3. Both t_r and t_{r+1} are independent since we get $t_r = X + Y$ and $t_{r+1} = N - x + W$. Then, event C is limited to three subpaths: $t_{r+1} = r + 1$ (Path 3-1), $N - x = r + 1 + x \wedge t_{r+1} = j_{r+1}$ (Path 3-2) and $t_{r+1} = t_r$ except when t_r equals either $r + 1$ or j_{r+1} (Path 3-3). We can compute the probability of each subpath as follows.

$$\begin{aligned}
\Pr(\text{Path 3-1}) &= \Pr(\text{Path 3} \wedge t_{r+1} = r + 1) = 1,\\
\Pr(\text{Path 3-2}) &= \Pr(\text{Path 3} \wedge N - x = r + 1 + x \wedge t_{r+1} = j_{r+1}) = 1,\\
\Pr(\text{Path 3-3}) &= \Pr(\text{Path 3} \wedge t_{r+1} = t_r) = 1 - \tfrac{2}{N}.
\end{aligned}$$

Therefore, the probability of Path 3 is computed as follows.

$$\begin{aligned}
\Pr(\text{Path 3}) = {} & \Pr(\text{Path 3-1}) \cdot \Pr(t_{r+1} = r + 1)\\
& + \Pr(\text{Path 3-2}) \cdot \Pr(N - x = r + 1 + x \wedge t_{r+1} = j_{r+1})\\
& + \Pr(\text{Path 3-3}) \cdot \Pr(t_{r+1} = t_r)\\
\approx {} & 1 \cdot \frac{1}{N} + 1 \cdot \left(\frac{1}{N} \cdot \frac{1}{N}\right) + \left(1 - \frac{2}{N}\right) \cdot \frac{1}{N} = \frac{1}{N}\left(2 - \frac{1}{N}\right).
\end{aligned}$$

From these results, $\Pr(C|B' \wedge A)$ is computed as follows.

$$\begin{aligned}
\Pr(C|B' \wedge A) = {} & \Pr(\text{Path 1}) \cdot \Pr(j_r = r) + \Pr(\text{Path 2}) \cdot \Pr(j_r = r + 1)\\
& + \Pr(\text{Path 3}) \cdot \Pr(j_r \neq r, r + 1)\\
\approx {} & \frac{1}{N}\left(2 - \frac{1}{N}\right) \cdot \frac{1}{N} + \frac{1}{N}\left(1 + \frac{1}{N}\right) \cdot \frac{1}{N} + \frac{1}{N}\left(2 - \frac{1}{N}\right) \cdot \left(1 - \frac{2}{N}\right)\\
= {} & \frac{2}{N}\left(1 - \frac{1}{N} + \frac{1}{N^2}\right).
\end{aligned}$$

As a result, $\Pr(A|B \wedge C)$ is computed as follows.

$$\Pr(A|B \wedge C) \approx \Pr(C|B' \wedge A) \approx \frac{2}{N}\left(1 - \frac{1}{N} + \frac{1}{N^2}\right). \qquad \square$$

Finally, we can integrate long-term Glimpse on $S_r[r + 1]$ as Theorem 7.

Theorem 7. *After the r-th round of PRGA for $r \geq 1$ and $\forall x \in [0, N - 1]$, we have*

$$\Pr(S_r[r + 1] = N - x | Z_{r+1} = Z_r \wedge Z_{r+1} = r + 1 + x) \approx \begin{cases} \frac{1}{N}\left(1 - \frac{2}{N^2}\right) & \text{if } x = 0\\ \frac{1}{N}\left(3 - \frac{6}{N} + \frac{2}{N^2}\right) & \text{if } x = 1^{1}\\ \frac{2}{N}\left(1 - \frac{1}{N} + \frac{1}{N^2}\right) & \text{otherwise.} \end{cases}$$

[1] The probability of correlation when $x = 1$ can be precisely revised to $\frac{1}{N}(3 - \frac{6}{N} + \frac{2}{N^2})$ from [7] in the same way as our other cases of $x \neq 1$, whose precise proof will be given in the final paper.

5 Experimental results

In order to check the accuracy of biases shown in Theorems 4 to 6, the experiments are executed using 2^{24} randomly chosen keys of 16 bytes and 2^{24} output keystreams for each key, which mean $2^{48}(= N^6)$ trials of RC4. Note that $\mathcal{O}(N^3)$ trials are reported to be sufficient to identify the biases with reliable success probability since each correlation here is of about $\frac{1}{N}$ with respect to a base event of probability $\frac{1}{N}$. Our experimental environment is as follows: Linux machine with 2.6 GHz CPU, 3.8 GiB memory, gcc 4.6.3 compiler and C language. We also evaluate the percentage of relative error ϵ of experimental values compared with theoretical values:

$$\epsilon = \frac{|\text{experimental value} - \text{theoretical value}|}{\text{experimental value}} \times 100(\%).$$

Table 1. Comparison between experimental and theoretical values

Results		Experimental value	Theoretical value	$\epsilon(\%)$
Theorem 4		0.000030522	0.000030398	0.406
Theorem 5	for $x = 1$	0.003922408	0.003906131	0.415
	for $x = 255$	0.000030683	0.000030398	0.929
	for $x = N - r$ $(x \neq 1, 255)$	0.000015259	0.000015140	0.780
Theorem 6		0.007812333	0.007782102	0.387

Table 1 shows experimental, theoretical values and the percentage of relative errors ϵ, which indicates ϵ is small enough in each case such as $\epsilon \leq 0.929$. Therefore, we have convinced that theoretical values closely reflects the experimental values.

6 Conclusion

In this paper, we have shown dual cases of the previous long-term Glimpse. We have also shown a new long-term Glimpse. We note that the previous long-term Glimpse is limited to $S_r[r+1] = N-1$ but that our results varies $S_r[r+1] \in [0, N-2]$. As a result, these long-term Glimpse can be integrated to biases of $S_r[r+1] \in [0, N-1]$. These new integrated long-term Glimpse could contribute to the improvement of state recovery attack on RC4, which remains an open problem.

References

1. Apurba Das, Subhamoy Maitra, Goutam Paul, and Santanu Sarkar. Some combinatorial results towards state recovery attack on rc4. In Sushil Jajodia and Chandan Mazumdar, editors, *Information Systems Security - ICISS 2011*, volume 7093

of *Lecture Notes in Computer Science*, pages 204–214. Springer Berlin Heidelberg, 2011.

2. Sourav Sen Gupta, Subhamoy Maitra, Goutam Paul, and Santanu Sarkar. Proof of empirical rc4 biases and new key correlations. In Ali Miri and Serge Vaudenay, editors, *Selected Areas in Cryptography - SAC 2011*, volume 7118 of *Lecture Notes in Computer Science*, pages 151–168. Springer Berlin Heidelberg, 2011.
3. Sourav Sen Gupta, Subhamoy Maitra, Goutam Paul, and Santanu Sarkar. (non-)random sequences from (non-)random permutations - analysis of rc4 stream cipher. *Journal of Cryptology*, 27(1):67–108, 2014.
4. Takanori Isobe, Toshihiro Ohigashi, Yuhei Watanabe, and Masakatu Morii. Full plaintext recovery attack on broadcast rc4. In *Fast Software Encryption - FSE 2013*, volume 8424 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013.
5. R. J. Jenkins. Isaac and rc4, 1996.
6. Lars R. Knudsen, Willi Meier, Bart Preneel, Vincent Rijmen, and Sven Verdoolaege. Analysis methods for (alleged) rc4. In Kazuo Ohta and Dingyi Pei, editors, *Advances in Cryptology - ASIACRYPT '98*, volume 1514 of *Lecture Notes in Computer Science*, pages 327–341. Springer Berlin Heidelberg, 1998.
7. Subhamoy Maitra and Sourav Sen Gupta. New long-term glimpse of rc4 stream cipher. In Aditya Bagchi and Indrakshi Ray, editors, *Information Systems Security - ICISS 2013*, volume 8303 of *Lecture Notes in Computer Science*, pages 345–359. Springer Berlin Heidelberg, 2013.
8. Subhamoy Maitra, Goutam Paul, Santanu Sarkar, Michael Lehmann, and Willi Meier. New results on generalization of roos-type biases and related keystreams of rc4. In Amr Youssef, Abderrahmane Nitaj, and Aboul Ella Hassanien, editors, *Progress in Cryptology - AFRICACRYPT 2013*, volume 7918 of *Lecture Notes in Computer Science*, pages 222–239. Springer Berlin Heidelberg, 2013.
9. Alexander Maximov and Dmitry Khovratovich. New state recovery attack on rc4. In David Wagner, editor, *Advances in Cryptology - CRYPTO 2008*, volume 5157 of *Lecture Notes in Computer Science*, pages 297–316. Springer Berlin Heidelberg, 2008.
10. Santanu Sarkar, Sourav Sen Gupta, Goutam Paul, and Subhamoy Maitra. Proving tls-attack related open biases of rc4. *IACR Cryptology ePrint Archive*, 2013:502, 2013.
11. Pouyan Sepehrdad, Serge Vaudenay, and Martin Vuagnoux. Discovery and exploitation of new biases in rc4. In Alex Biryukov, Guang Gong, and Douglas R. Stinson, editors, *Selected Areas in Cryptography - SAC 2010*, volume 6544 of *Lecture Notes in Computer Science*, pages 74–91. Springer Berlin Heidelberg, 2011.