Many Keystream Bytes of RC4 Leak Secret Key Information

Subhamoy Maitra*, Goutam Paul†

Abstract

In this paper, we show that RC4 keystream leaks secret key information in the first 32 bytes and also in the 256-th and 257-th bytes. For the first time a complete framework is presented to show that many keystream output bytes are found to be significantly biased towards several linear combinations of the secret key bytes, without assuming any condition on the secret key.

Keywords: Bias, Cryptanalysis, Keystream, RC4, Stream Cipher.

1 Introduction

RC4 is one of the most well known stream ciphers. It has very simple implementation and it is being used in a number of commercial products till date. Being one of the popular stream ciphers, it has also been subjected to many cryptanalytic attempts for more than a decade. Though lots of weaknesses have already been explored in RC4 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18], it could not be thoroughly cracked yet and proper use of this stream cipher is still believed to be quite secure.

The Key Scheduling Algorithm (KSA) and the Pseudo Random Generation Algorithm (PRGA) of RC4 are presented below. The data structure contains an array S of size N (typically, 256), which contains a permutation of the integers $\{0, \ldots, N-1\}$, two indices i, j and the secret key array K. Given a secret key k of l bytes (typically 5 to 16), the array K of size N is such that $K[y] = k[y \bmod l]$ for any y, $0 \le y \le N-1$.

*Applied Statistics Unit, Indian Statistical Institute, 203 B T Road, Kolkata 700 108, India, Email: subho@isical.ac.in

†Department of Computer Science and Engineering, Jadavpur University, Kolkata 700 032, India, Email: goutam_paul@cse.jdvu.ac.in

Algorithm KSA	**Algorithm PRGA**
Initialization:	*Initialization*:
For $i = 0, \ldots, N-1$	$i = j = 0$;
$S[i] = i$;	*Output Keystream Generation Loop*:
$j = 0$;	$i = i + 1$;
Scrambling:	$j = j + S[i]$;
For $i = 0, \ldots, N-1$	Swap($S[i]$, $S[j]$);
$j = (j + S[i] + K[i])$;	$t = S[i] + S[j]$;
Swap($S[i], S[j]$);	Output $z = S[t]$;

Apart from some minor details, the KSA and the PRGA are almost same. In KSA, the updation of the index j depends on the secret key, whereas the key is not used in PRGA. One may consider the PRGA as the KSA with all zero key. All additions in both the KSA and the PRGA are additions modulo N.

There are two broad approaches in the study of cryptanalysis of RC4: attacks based on the weaknesses of the KSA and those based on the weaknesses of the PRGA. Distinguishing attacks are the main motivation for PRGA-based approach [2, 4, 7, 8, 9, 13, 14]. Important results in this approach include bias in the keystream output bytes. Initial works on distinguishing the RC4 keystream from random stream has been done in [4, 2]. A bias in the second output byte being zero has been proved in [7] and a bias in the equality of the first two output bytes has been shown in [14]. In [9], it has been shown that getting strings of pattern $ABGAB$ (A, B are bytes and G is a string of bytes of small length, say ≤ 16) are more probable in RC4 keystream than in random stream. In [10], RC4 has been analyzed using the theory of random shuffles and it has been recommended that initial 512 bytes of the keystream output should be discarded in order to be safe. Recently, differential attacks on RC4 have been discussed in [1, Section 5].

Initial empirical works based on the weaknesses of the RC4 KSA were done in [15, 18] and several classes of weak keys had been identified. In [15], experimental evidences have been reported that the first keystream output byte of RC4 leaks information about secret key when the first two secret key bytes add to 0 mod 256. Recently, a more general theoretical study has been performed in [11] which includes the observations of [15]. The work [11] shows how the bias of the "third permutation byte" (after the KSA) towards the "first three secret key bytes" propagates to the first keystream output byte (in the PRGA). This is proved for any secret key, i.e., there is no condition on the secret key bytes. The exact result depicts that the first keystream output byte is three more than the sum of first three secret key bytes with a probability $(1 + 0.37)\frac{1}{N}$. In [6], the biases in the initial bytes have been noted which we present in a concrete theoretical framework here. Very recently [12], it has been identified that if the permutation after the KSA, or the permutation at any stage of the PRGA with relevant information about the indices i, j are known, then the secret key bytes can be recovered efficiently.

The works [3, 8] also explain how secret key information is leaked in the keystream output bytes. In [3], it is considered that the first few bytes of the secret key is known (this is practical as in one mode of WEP the IV bytes and the secret key bytes are concatenated to get the effective key of RC4) and based on that the next byte of the secret key is predicted. The attack is based on how secret key information is leaked in the first keystream output byte of PRGA. In [8], the same idea of [3] has been exploited with the Glimpse theorem [5] to find the information leakage about the secret key at the 257-th byte of the PRGA. Later, the works [6, 16, 17] improve [3].

1.1 Our Contribution

Let S_r be the permutation after r many rounds of the RC4 KSA, $r \geq 1$. Hence S_N is the permutation after the complete key scheduling. By S_0, we denote the initial identity permutation. During the round r of the KSA, $r \geq 1$, the value of index i is $r-1$ and hence after the round r the permutation S_r can also be denoted by S_{i+1}.

Let S_r^G be the permutation, i_r and j_r be the values of the indices i and j, and z_r be the keystream output byte after r many rounds of the PRGA, $r \geq 1$. Clearly, $i_r = r \bmod N$. We also denote S_N by S_0^G as this is the permutation before the PRGA starts.

Further, let

$$f_y = \frac{y(y+1)}{2} + \sum_{x=0}^{y} K[x],$$

for $y \geq 0$. Note that all the additions and subtractions related to the key bytes are modulo N.

Our contribution can be summarized as follows.

- In Theorem 1 (Section 2.1), we theoretically prove that
  (1) $P(z_1 = 1 - f_1) = \frac{1}{N} \cdot \left(1 + (\frac{N-1}{N})^{N+2} + \frac{1}{N}\right)$, and
  (2) for $2 \leq r \leq N-1$, $P(z_r = r - f_r) =$
  $\frac{1}{N} \cdot \left(1 + \left[(\frac{N-r}{N}) \cdot (\frac{N-1}{N})^{[\frac{r(r+1)}{2}+N]} + \frac{1}{N}\right] \cdot \left[(\frac{N-1}{N})^{r-1} - \frac{1}{N}\right] + \frac{1}{N}\right)$.
  The bias is significant from $r = 1$ through $r = 32$, i.e., for the first 32 keystream output bytes of RC4.

  These biases have been identified in [6], but we present these results with detailed theoretical analysis.

- Using similar arguments, in Theorem 2 (Section 2.2), we show that $P(z_N = N - f_0) = \frac{1}{N} \cdot \left(1 + (\frac{N-1}{N})^{2N-1} + \frac{1}{N^2} \cdot (\frac{N-1}{N})^{N-1} - \frac{1}{N^2} + \frac{1}{N}\right)$. This indicates bias at z_{256}.

- Using the assumption $P(S_N[S_N[1]] = f_1) = (\frac{N-1}{N})^{2(N-1)}$, in Theorem 3 (Section 3), we prove $P(z_{N+1} = N + 1 - f_1) = \frac{1}{N} \cdot \left(1 + (\frac{N-1}{N})^{3(N-1)} - \frac{1}{N} \cdot (\frac{N-1}{N})^{2(N-1)} + \frac{1}{N}\right)$. This indicates bias at z_{257}. Further, we provide experimental justification towards $P(S_N[S_N[1]] = f_1) = (\frac{N-1}{N})^{2(N-1)}$ and related non-random associations between the permutation bytes and secret key bytes.

- In Section 4, we observe additional experimental biases in the initial keystream bytes showing $P(z_r = f_{r-1}) \geq \frac{1}{256} \cdot (1 + 0.05)$, for $r = 1$ to 21, except $r = 2$.

- Finally, we accumulate these results in Section 5 to present how one can guess the secret key from the keystream output bytes with non-negligible biases than random guess.

The works presented in [15, 3, 8, 6, 16, 17] assume certain conditions on the secret key bytes of RC4 in order to mount their attacks. We do not consider any such requirements here. Further, our work is much more general than [6, 11], as we show that there exist significant biases in many of the keystream output bytes (bytes 1 to 32 and also 256, 257) towards different linear combinations of secret key bytes.

2 New Biases in RC4 Keystream: Theoretical Results

We start with some existing related results.

Proposition 1 *[12] Consider that the index j takes its values uniformly at random during the KSA rounds. Then, $P(S_N[y] = f_y) = (\frac{N-y}{N}) \cdot (\frac{N-1}{N})^{[\frac{y(y+1)}{2}+N]} + \frac{1}{N}$, $0 \le y \le N-1$.*

As explained in [12], the above result indicates significant biases for small values of y (more precisely, for $0 \le y \le 47$), as is supported by the experimental result presented in [15].

The Glimpse Main Theorem [5, 8] states that after the r-th round of the PRGA, $r \ge 1$, $P(S_r^G[j_r] = r - z_r) = P(S_r^G[i_r] = j_r - z_r) = \frac{2}{N}$. We rewrite the first relation between $S_r^G[j_r]$ and $r - z_r$ as the following proposition.

Proposition 2 *$P(z_r = r - S_{r-1}^G[i_r]) = \frac{2}{N}$ for $r \ge 1$.*

Proof: $S_r^G[j_r]$ is assigned the value of $S_{r-1}^G[i_r]$. As the Glimpse Main Theorem gives $P(z_r = r - S_r^G[j_r]) = \frac{2}{N}$ for $r \ge 1$, we get $P(z_r = r - S_{r-1}^G[i_r]) = \frac{2}{N}$ for $r \ge 1$. ■

The idea of writing the Glimpse Main Theorem in the form of Proposition 2 is due to the fact that relating "z_r to $S_{r-1}^G[i_r]$" will ultimately relate "z_r with secret key bytes" as "initial values of the permutations in initial rounds of PRGA" are related to "secret key".

Now we start with our results. The following lemma shows how the permutation bytes at rounds t and $r-1$ of the PRGA, for $t+2 \le r$, are related.

Lemma 1 *Let $P(S_t^G[i_r] = X) = q_{t,r}$, for any value X. Then, for $t+2 \le r \le t+N$, $P(S_{r-1}^G[i_r] = X) = q_{t,r} \cdot \left[(\frac{N-1}{N})^{r-t-1} - \frac{1}{N}\right] + \frac{1}{N}$.*

Proof: We consider two separate cases.

1. $S_t^G[i_r] = X$ and during the next $(r-t-1)$ rounds of the PRGA, the index i_r is not touched by any of the $r-t-1$ many j values (since $t+2 \le r \le t+N$, the index i_r is not touched by any of the $r-t-1$ many i values anyway). The first event occurs with probability $q_{t,r}$ and the second event occurs with probability $(\frac{N-1}{N})^{r-t-1}$. Thus the contribution of this case is $q_{t,r} \cdot (\frac{N-1}{N})^{r-t-1}$.

2. $S_t^G[i_r] \neq X$ and still $S_{r-1}^G[i_r]$ equals X by random association. The contribution of this case is $(1 - q_{t,r}) \cdot \frac{1}{N}$.

Thus, adding the above two contributions, we get $P(S_{r-1}^G[i_r] = X) = q_{t,r} \cdot (\frac{N-1}{N})^{r-t-1} + (1 - q_{t,r}) \cdot \frac{1}{N} = q_{t,r} \cdot \left[(\frac{N-1}{N})^{r-t-1} - \frac{1}{N}\right] + \frac{1}{N}$. ■

Note that the above result holds for $t+2 \leq r \leq t+N$, and not for $r = t+1$. If we take $r = t+1$, then $S_{r-1}^G = S_t^G$, which is our starting point, i.e., $P(S_{r-1}^G[i_r] = X) = P(S_t^G[i_r] = X) = q_{t,r}$.

Next, we present the bias of each keystream output byte to a combination of the secret key bytes in the following lemma.

Lemma 2 *Let $P(S_{r-1}^G[i_r] = f_{i_r}) = w_r$, for $r \geq 1$. Then $P(z_r = r - f_{i_r}) = \frac{1}{N} \cdot (1 + w_r)$.*

Proof: We consider two separate cases in which the event $(z_r = r - f_{i_r})$ can occur.

1. $S_{r-1}^G[i_r] = f_{i_r}$ and $z_r = r - S_{r-1}^G[i_r]$. The contribution of this case is $P(S_{r-1}^G[i_r] = f_{i_r}) \cdot P(z_r = r - S_{r-1}^G[i_r]) = w_r \cdot \frac{2}{N}$ (by Proposition 2).

2. $S_{r-1}^G[i_r] \neq f_{i_r}$, and still $z_r = r - f_{i_r}$ due to random association. So the contribution of this case is $P(S_{r-1}^G[i_r] \neq f_{i_r}) \cdot \frac{1}{N} = (1 - w_r) \cdot \frac{1}{N}$.

Adding the above two contributions, we get the total probability as $w_r \cdot \frac{2}{N} + (1 - w_r) \cdot \frac{1}{N} = \frac{1}{N} \cdot (1 + w_r)$. ■

2.1 Biases in the initial keystream output bytes

The results in this section for biases in initial keystream bytes has earlier been pointed out in [6]. However, the exact theoretical formulae for the biases of the different keystream output bytes has not been attempted in [6].

Theorem 1
(1) $P(z_1 = 1 - f_1) = \frac{1}{N} \cdot \left(1 + (\frac{N-1}{N})^{N+2} + \frac{1}{N}\right)$.
(2) *For* $2 \leq r \leq N-1$,
$P(z_r = r - f_r) = \frac{1}{N} \cdot \left(1 + \left[(\frac{N-r}{N}) \cdot (\frac{N-1}{N})^{[\frac{r(r+1)}{2}+N]} + \frac{1}{N}\right] \cdot \left[(\frac{N-1}{N})^{r-1} - \frac{1}{N}\right] + \frac{1}{N}\right)$.

Proof: First, we prove part (1). In the first round, i.e., when $r = 1$, $i_r = 1$ and $f_{i_r} = f_1$, and so $w_1 = P(S_0^G[1] = f_1) = P(S_N[1] = f_1) = (\frac{N-1}{N}) \cdot (\frac{N-1}{N})^{[\frac{1(1+1)}{2}+N]} + \frac{1}{N} = (\frac{N-1}{N})^{N+2} + \frac{1}{N}$ (by Proposition 1). Now, using Lemma 2, we get $P(z_1 = 1 - f_1) = \frac{1}{N} \cdot (1 + w_1) = \frac{1}{N} \cdot \left(1 + (\frac{N-1}{N})^{N+2} + \frac{1}{N}\right)$.

Next, we prove part (2). For $2 \leq r \leq N-1$, we have $i_r = r$ and $f_{i_r} = f_r$. Taking $X = f_r$ and $t = 0$ in Lemma 1, we have $q_{0,r} = P(S_0^G[r] = f_r) = P(S_N[r] = f_r) = (\frac{N-r}{N}) \cdot (\frac{N-1}{N})^{[\frac{r(r+1)}{2}+N]} + \frac{1}{N}$ (by Proposition 1), and hence $w_r = P(S_{r-1}^G[r] = f_r) = \left[(\frac{N-r}{N}) \cdot (\frac{N-1}{N})^{[\frac{r(r+1)}{2}+N]} + \frac{1}{N}\right] \cdot \left[(\frac{N-1}{N})^{r-1} - \frac{1}{N}\right] + \frac{1}{N}$. Now, using Lemma 2, we get $P(z_r = r - f_r) = \frac{1}{N} \cdot (1 + w_r) = \frac{1}{N} \cdot \left(1 + \left[(\frac{N-r}{N}) \cdot (\frac{N-1}{N})^{[\frac{r(r+1)}{2}+N]} + \frac{1}{N}\right] \cdot \left[(\frac{N-1}{N})^{r-1} - \frac{1}{N}\right] + \frac{1}{N}\right)$. ■

Note that Lemma 1 is not used in proving part (1) of the above theorem. It is proved directly from Proposition 1. In fact, Lemma 1 can not be used in part (1), as here we have $r = t + 1$ with $t = 0$.

To have a clear understanding of the quantity of the biases, Table 1 lists the numerical values of the probabilities according to the formula given in Theorem 1. Note that the random association is $\frac{1}{N}$, which is 0.0039 for $N = 256$.

Close to the round 48, the biases tend to disappear. This is indicated by the convergence of the values to the probability $\frac{1}{256} = 0.0039$.

r	$P(z_r = r - f_r)$							
1-8	0.0053	0.0053	0.0053	0.0053	0.0052	0.0052	0.0052	0.0051
9-16	0.0051	0.0050	0.0050	0.0049	0.0048	0.0048	0.0047	0.0047
17-24	0.0046	0.0046	0.0045	0.0045	0.0044	0.0044	0.0043	0.0043
25-32	0.0043	0.0042	0.0042	0.0042	0.0041	0.0041	0.0041	0.0041
33-40	0.0041	0.0040	0.0040	0.0040	0.0040	0.0040	0.0040	0.0040
41-48	0.0040	0.0040	0.0040	0.0040	0.0040	0.0039	0.0039	0.0039

Table 1: The probabilities computed following Theorem 1.

One may check that $P(z_1 = 1 - f_1) = \frac{1}{N}(1 + 0.36)$ and that decreases to $P(z_{32} = 32 - f_{32}) = \frac{1}{N}(1 + 0.05)$, but still then it is 5% more than the random association. These associations have also been pointed out in [6] in relation to WEP attacks. Our results are based on more rigorous theoretical analysis than [6].

2.2 Bias in the 256-th keystream output byte

Interestingly, the bias again reappears after round 256 and round 257. First we present a bias for the 256-th keystream byte.

Theorem 2 $P(z_N = N - f_0) = \frac{1}{N} \cdot \left(1 + (\frac{N-1}{N})^{2N-1} + \frac{1}{N^2} \cdot (\frac{N-1}{N})^{N-1} - \frac{1}{N^2} + \frac{1}{N}\right)$.

Proof: During the N-th round of the PRGA, $i_N = N \bmod N = 0$. Taking $X = f_0$, $t = 0$ and $r = N$ in Lemma 1, we have $q_{0,N} = P(S_0^G[0] = f_0) = P(S_N[0] = f_0) = (\frac{N-1}{N})^N + \frac{1}{N}$ (by Proposition 1), and hence $w_N = P(S_{N-1}^G[0] = f_0) = \left[(\frac{N-1}{N})^N + \frac{1}{N}\right] \cdot \left[(\frac{N-1}{N})^{N-1} - \frac{1}{N}\right] + \frac{1}{N} = (\frac{N-1}{N})^{2N-1} + \frac{1}{N^2} \cdot (\frac{N-1}{N})^{N-1} - \frac{1}{N^2} + \frac{1}{N}$. Thus, by Lemma 2, the bias is given by $P(z_N = N - f_0) = \frac{1}{N} \cdot (1 + w_N) = \frac{1}{N} \cdot \left(1 + (\frac{N-1}{N})^{2N-1} + \frac{1}{N^2} \cdot (\frac{N-1}{N})^{N-1} - \frac{1}{N^2} + \frac{1}{N}\right)$. ■

For $N = 256$, $w_N = w_{256} = 0.1392$ and the bias turns out to be 0.0045 (i.e., $\frac{1}{256}(1 + 0.1392)$). Experimental results confirm all the theoretical values presented in this section.

3 Bias in the 257-th keystream output byte

For the bias on the 257-th output byte, we depend on the experimental observation that $P(S_N[S_N[1]] = f_1) = (\frac{N-1}{N})^{2(N-1)}$. We could not theoretically prove this observation so far. Before going for more discussion on this observation, let us first assume this result and prove how one can get a bias in the 257-th keystream output byte.

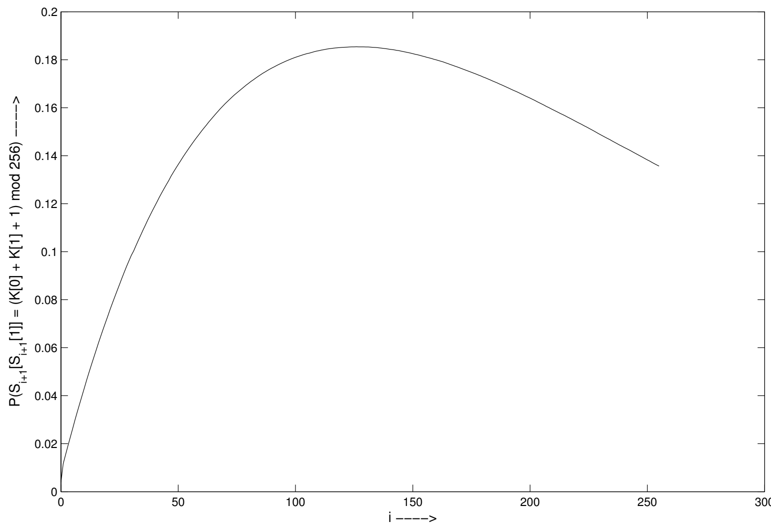

Figure 1: $P(S_{i+1}[S_{i+1}[1]] = f_1)$ versus i $(r = i+1)$ during RC4 KSA.

Theorem 3 *Given* $P(S_N[S_N[1]] = f_1) = (\frac{N-1}{N})^{2(N-1)}$,

$$P(z_{N+1} = N+1-f_1) = \frac{1}{N} \cdot \Big(1 + (\frac{N-1}{N})^{3(N-1)} - \frac{1}{N} \cdot (\frac{N-1}{N})^{2(N-1)} + \frac{1}{N}\Big).$$

Proof: During the $(N+1)$-th round, we have, $i_{N+1} = (N+1) \bmod N = 1$. Taking $X = f_1$, $t = 1$ and $r = N+1$ in Lemma 1, we have $q_{1,N+1} = P(S_1^G[1] = f_1) = P(S_N[S_N[1]] = f_1) = (\frac{N-1}{N})^{2(N-1)}$, and hence $w_{N+1} = P(S_N^G[1] = f_1) = (\frac{N-1}{N})^{2(N-1)} \cdot \big[(\frac{N-1}{N})^{N-1} - \frac{1}{N}\big] + \frac{1}{N} = (\frac{N-1}{N})^{3(N-1)} - \frac{1}{N} \cdot (\frac{N-1}{N})^{2(N-1)} + \frac{1}{N}$. Now, using Lemma 2, we get $P(z_{N+1} = N+1-f_1) = \frac{1}{N} \cdot (1 + w_{N+1}) = \frac{1}{N} \cdot \Big(1 + (\frac{N-1}{N})^{3(N-1)} - \frac{1}{N} \cdot (\frac{N-1}{N})^{2(N-1)} + \frac{1}{N}\Big)$. ■

For $N = 256$, $w_{N+1} = w_{257} = 0.0535$ and $P(z_{257} = 257 - f_1) = \frac{1}{N} \cdot (1 + 0.0535) = 0.0041$ which also conforms to experimental observation.

At this point we like to point out how $P(S_r[S_r[1]] = f_1)$ varies with the rounds $r, 1 \leq r \leq N$, of the KSA of RC4. Once again, note that $f_1 = (K[0] + K[1] + 1) \bmod N$. We experimented on 10 million randomly chosen secret keys. Figure 1 demonstrates the experimental results that $P(S_r[S_r[1]] = f_1)$ increases till around $r = \frac{N}{2}$ where it gets the maximum value around 0.185 and then it decreases to 0.136 at $r = N$. On the other hand, the value of the expression $(\frac{N-1}{N})^{2(N-1)}$ is also 0.136 for $N = 256$ and that is why, based on experimental observation, we have used the assumption $P(S_N[S_N[1]] = f_1) = (\frac{N-1}{N})^{2(N-1)}$ in Theorem 3.

Though we could not prove the result $P(S_N[S_N[1]] = f_1) = (\frac{N-1}{N})^{2(N-1)}$ theoretically, we could prove this for the case $r = 2$, i.e., after the round 2 of RC4 KSA.

Proposition 3 $P(S_2[S_2[1]] = K[0] + K[1] + 1) = \frac{3}{N} - \frac{4}{N^2} + \frac{2}{N^3} = \frac{3}{N}$.

Proof: The proof is based on three cases.

1. Let $K[0] \neq 0, K[1] = N - 1$. The probability of this event is $\frac{N-1}{N^2}$. Now $S_2[1] = S_1[K[0] + K[1] + 1] = S_1[K[0]] = S_0[0] = 0$. So, $S_2[S_2[1]] = S_2[0] = S_1[0] = K[0] = K[0] + K[1] + 1$. Note that $S_2[0] = S_1[0]$, as $K[0] + K[1] + 1 \neq 0$.

2. Let $K[0] + K[1] = 0, K[0] \neq 1$, i.e., $K[1] \neq N - 1$. The probability of this event is $\frac{N-1}{N^2}$. Now $S_2[1] = S_1[K[0] + K[1] + 1] = S_1[1] = S_0[1] = 1$. Note that $S_1[1] = S_0[1]$, as $K[0] \neq 1$. So, $S_2[S_2[1]] = S_2[1] = 1 = K[0] + K[1] + 1$.

3. $S_2[S_2[1]]$ could be $K[0] + K[1] + 1$ by random association except the two previous cases.

Thus $P(S_2[S_2[1]] = K[0] + K[1] + 1) = \frac{2(N-1)}{N^2} + (1 - \frac{2(N-1)}{N^2})\frac{1}{N} = \frac{3}{N} - \frac{4}{N^2} + \frac{2}{N^3}$. ■

The theoretical value of the expression $\frac{3}{N} - \frac{4}{N^2} + \frac{2}{N^3}$ for $N = 256$ is 0.011658 and it matches with experimental observation.

Proposition 3 shows that after the second round ($i = 1, r = 2$), the event ($S_2[S_2[1]] = K[0] + K[1] + 1$) is not a random association.

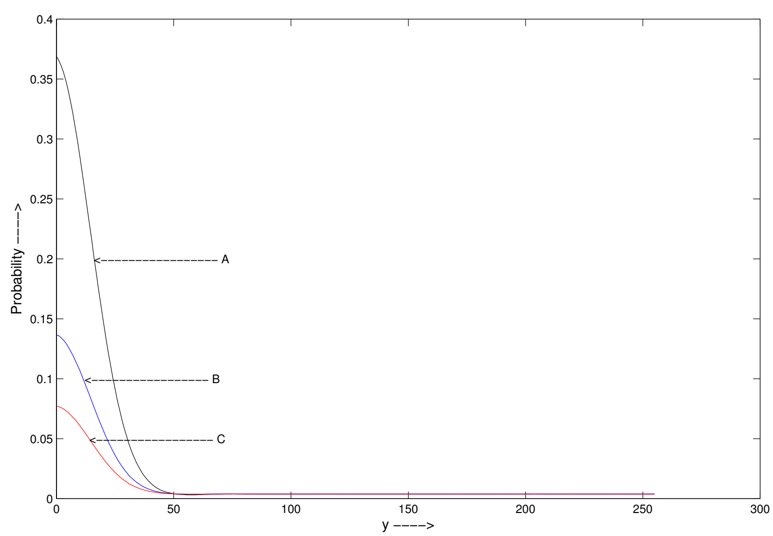

Figure 2: A: $P(S_N[y] = f_y)$, B: $P(S_N[S_N[y]] = f_y)$, C: $P(S_N[S_N[S_N[y]]] = f_y)$ versus y for $0 \leq y \leq 255$.

Now we like to present a more detailed observation. In [15, 12], the association between $S_N[y]$ and f_y is shown. As we have observed the non-random association between $S_N[S_N[1]]$ and f_1, it is important to study what is the association between $S_N[S_N[y]]$ and f_y, and moving further, the association between $S_N[S_N[S_N[y]]]$ and f_y, for $0 \leq y \leq N-1$. Our experimental observations show that all these associations are not random (i.e., much more than $\frac{1}{N}$) for initial values of y. The experimental observations (over 10 million runs of randomly chosen keys) are presented in Figure 2 and also in the Table 3 in Appendix A. It will be of great interest to theoretically study the association between $S_N[S_N \ldots [S_N[y]] \ldots]$ and f_y in general.

4 Further Biases in RC4 Keystream: Experimental Observation

We also experimentally observe some other significant biases that we could not prove theoretically so far. One may easily simulate the experiments to check our claims.

r	$P(z_r = f_{r-1})$
1-8	0.0043 0.0039 0.0044 0.0044 0.0044 0.0044 0.0043 0.0043
9-16	0.0043 0.0043 0.0043 0.0042 0.0042 0.0042 0.0042 0.0042
17-24	0.0041 0.0041 0.0041 0.0041 0.0041 0.0040 0.0040 0.0040
25-32	0.0040 0.0040 0.0040 0.0040 0.0040 0.0040 0.0040 0.0040
33-40	0.0039 0.0039 0.0039 0.0039 0.0039 0.0039 0.0039 0.0039
41-48	0.0039 0.0039 0.0039 0.0039 0.0039 0.0039 0.0039 0.0039

Table 2: Additional bias of the keystream bytes towards the secret key.

The results are presented in Table 2 which is experimented over hundred million (10^8) randomly chosen keys of 16 bytes. For proper random association, $P(z_r = f_{r-1})$ should have been $\frac{1}{256}$, i.e., 0.0039. However, this is not true for RC4 keystream generation and experimental results show that $P(z_r = f_{r-1}) \geq \frac{1}{256}(1 + 0.05)$ for $1 \leq r \leq 21$ except $r = 2$.

5 Cryptanalytic Applications

Here we accumulate the theoretical and experimental results of the previous two sections. Consider the first keystream output byte z_1 of PRGA. We find the theoretical result that $P(z_1 = 1 - f_1) = 0.0053$ (see Theorem 1 and Table 1) and the experimental observation that $P(z_1 = f_0) = 0.0043$ (see Table 2). Further, from [11], we have the result that $P(z_1 = f_2) = 0.0053$. Taking them together, one can check that the $P(z_1 = f_0 \vee z_1 = 1 - f_1 \vee z_1 = f_2) = 1-(1-0.0043)\cdot(1-0.0053)\cdot(1-0.0053) = 0.0148$. Our result indicates that out of randomly chosen 10000 secret keys, in 148 cases on an average, z_1 reveals f_0 or $1 - f_1$ or f_2, i.e., $K[0]$ or $1-(K[0]+K[1]+1)$ or $(K[0]+K[1]+K[2]+3)$. If, however, one tries a random association, considering that z_1 will be among three randomly chosen values $\alpha_1, \alpha_2, \alpha_3$ from $[0, \ldots, 255]$, then $P(z_1 = \alpha_1 \vee z_1 = \alpha_2 \vee z_1 = \alpha_3) = 1 - (1 - \frac{1}{256})^3 = 0.0117$.

Thus one can guess z_1 with an additional advantage of $\frac{0.0148-0.0117}{0.0117} \cdot 100\% = 27\%$ over the random guess.

Now consider the keystream output byte z_2. We have $P(z_2 = 2 - f_2) = 0.0053$ (see Theorem 1 and Table 1), which provides an advantage of $\frac{0.0053-0.0039}{0.0039} \cdot 100\% = 36\%$.

Similarly, referring to Table 1 and Table 2, significant biases can be observed in $P(z_r = f_{r-1} \vee z_r = r - f_r)$ for $r = 3$ to 32 over random association.

Now consider the following scenario with the events $E_1, \ldots, E_{32}$, where $E_1 : (z_1 = f_0 \vee z_1 = 1 - f_1 \vee z_1 = f_2)$, $E_2 : (z_2 = 2 - f_2)$, and $E_r : (z_r = f_{r-1} \vee z_r = r - f_r)$ for $3 \leq r \leq 32$. Observing the first 32 keystream output bytes $z_1, \ldots, z_{32}$, one may try to guess the secret key assuming that 3 or more of the events $E_1, \ldots, E_{32}$ occur. We experimented with 10 million randomly chosen secret keys of length 16 bytes. We found that 3 or more of the events occur in 0.0028 proportion of cases, which is true for 0.0020 proportion of cases for random association. This demonstrates a substantial advantage (40%) over random guess.

6 Conclusion

In this paper we present several new observations on weakness of RC4. We present theoretical as well as experimental biases of the keystream output bytes towards the linear combinations of secret key bytes. Theoretical results are proved to show that RC4 keystream output bytes, at the indices 1 to 32 and then at 256, leak significant information about secret key bytes. Experimental observations and theoretical results are combined to identify that the 257-th keystream output byte is biased too towards secret key bytes. Further biases (apart from our theoretical results) of the initial keystream bytes have also been observed experimentally. This is the first time, many biases of the keystream output bytes of RC4 are discovered without any assumption on secret keys.

References

[1] E. Biham and O. Dunkelman. Differential Cryptanalysis in Stream Ciphers. IACR Eprint Server, eprint.iacr.org, number 2007/218, June 6, 2007.

[2] S. R. Fluhrer and D. A. McGrew. Statistical Analysis of the Alleged RC4 Keystream Generator. FSE 2000, pages 19-30, vol. 1978, Lecture Notes in Computer Science, Springer-Verlag.

[3] S. R. Fluhrer, I. Mantin and A. Shamir. Weaknesses in the Key Scheduling Algorithm of RC4. Selected Areas in Cryptography 2001, pages 1-24, vol. 2259, Lecture Notes in Computer Science, Springer-Verlag.

[4] J. Golic. Linear statistical weakness of alleged RC4 keystream generator. EUROCRYPT 1997, pages 226-238, vol. 1233, Lecture Notes in Computer Science, Springer-Verlag.

[5] R. J. Jenkins. ISAAC and RC4. 1996
Available at `http://burtleburtle.net/bob/rand/isaac.html`.

[6] A. Klein. Attacks on the RC4 stream cipher. February 27, 2006.
Available at `http://cage.ugent.be/ klein/RC4/`, [last accessed on June 27, 2007].

[7] I. Mantin and A. Shamir. A Practical Attack on Broadcast RC4. FSE 2001, pages 152-164, vol. 2355, Lecture Notes in Computer Science, Springer-Verlag.

[8] I. Mantin. A Practical Attack on the Fixed RC4 in the WEP Mode. ASIACRYPT 2005, pages 395-411, vol. 3788, Lecture Notes in Computer Science, Springer-Verlag.

[9] I. Mantin. Predicting and Distinguishing Attacks on RC4 Keystream Generator. EUROCRYPT 2005, pages 491-506, vol. 3494, Lecture Notes in Computer Science, Springer-Verlag.

[10] I. Mironov. (Not So) Random Shuffles of RC4. CRYPTO 2002, pages 304-319, vol. 2442, Lecture Notes in Computer Science, Springer-Verlag.

[11] G. Paul, S. Rathi and S. Maitra. On Non-negligible Bias of the First Output Byte of RC4 towards the First Three Bytes of the Secret Key. Proceedings of the International Workshop on Coding and Cryptography 2007, pages 285-294.

[12] G. Paul and S. Maitra. Permutation after RC4 Key Scheduling Reveals the Secret Key. In *14th Annual Workshop on Selected Areas in Cryptography*, SAC 2007, August 16-17, Ottawa, Canada.

[13] S. Paul and B. Preneel. Analysis of Non-fortuitous Predictive States of the RC4 Keystream Generator. INDOCRYPT 2003, pages 52-67, vol. 2904, Lecture Notes in Computer Science, Springer-Verlag.

[14] S. Paul and B. Preneel. A New Weakness in the RC4 Keystream Generator and an Approach to Improve the Security of the Cipher. FSE 2004, pages 245-259, vol. 3017, Lecture Notes in Computer Science, Springer-Verlag.

[15] A. Roos. A class of weak keys in the RC4 stream cipher. Two posts in sci.crypt, message-id `43u1eh$1j3@hermes.is.co.za` and `44ebge$llf@hermes.is.co.za`, 1995. Available at `http://marcel.wanda.ch/Archive/WeakKeys`.

[16] E. Tews, R. -P. Weinmann and Andrei Pyshkin. Breaking 104 bit WEP in less than 60 seconds. IACR Eprint Server, eprint.iacr.org, number 2007/120, April 1, 2007.

[17] S. Vaudenay and M. Vuagnoux. Passive-only key recovery attacks on RC4. In *14th Annual Workshop on Selected Areas in Cryptography*, SAC 2007, August 16-17, Ottawa, Canada.

[18] D. Wagner. My RC4 weak keys.
Post in sci.crypt, message-id 447o1l$cbj@cnn.Princeton.EDU, 26 September, 1995.
Available at http://www.cs.berkeley.edu/~daw/my-posts/my-rc4-weak-keys.

Appendix A

Table 3 starts here.

$S_N^2[y]$ denotes $S_N[S_N[y]]$ and $S_N^3[y]$ denotes $S_N[S_N[S_N[y]]]$.

y	$P(S_N[y] = f_y)$	$P(S_N^2[y] = f_y)$	$P(S_N^3[y] = f_y)$
0	0.368798	0.136546	0.077260
1	0.365703	0.135954	0.076768
2	0.361241	0.134267	0.075963
3	0.355746	0.132284	0.074879
4	0.348770	0.129842	0.073393
5	0.340758	0.126786	0.071970
6	0.331630	0.123272	0.070008
7	0.321588	0.119712	0.067876
8	0.309867	0.115536	0.065870
9	0.298392	0.111257	0.063389
10	0.285270	0.106890	0.060884
11	0.272321	0.102029	0.058212
12	0.258478	0.096909	0.055537
13	0.244201	0.091964	0.052644
14	0.230242	0.086878	0.049807
15	0.216420	0.081617	0.046952
16	0.201842	0.076551	0.044245
17	0.188238	0.071439	0.041333
18	0.174166	0.066379	0.038645
19	0.161092	0.061533	0.035903
20	0.147733	0.056873	0.033326
21	0.135591	0.052242	0.030786
22	0.123264	0.048053	0.028409
23	0.112470	0.043901	0.026082
24	0.101382	0.039950	0.024028
25	0.091580	0.036257	0.021883
26	0.081977	0.032895	0.020063
27	0.073780	0.029647	0.018269
28	0.065305	0.026827	0.016668
29	0.058091	0.024006	0.015181
30	0.051582	0.021659	0.013812
31	0.045121	0.019018	0.012293
32	0.039750	0.017318	0.011413
33	0.034991	0.015610	0.010443
34	0.030508	0.013937	0.009541
35	0.026730	0.012559	0.008747
36	0.023253	0.011194	0.008047
37	0.020120	0.010041	0.007394
38	0.017657	0.009123	0.006823
39	0.015299	0.008254	0.006420
40	0.013253	0.007514	0.005926
41	0.011550	0.006865	0.005596
42	0.010014	0.006319	0.005344
43	0.008733	0.005795	0.005007
44	0.007619	0.005442	0.004764
45	0.006732	0.005061	0.004550
46	0.006035	0.004771	0.004428
47	0.005358	0.004524	0.004263
48	0.004958	0.004357	0.004169
49	0.004467	0.004169	0.004065
50	0.004121	0.004007	0.003992
51	0.003860	0.003945	0.003924
52	0.003629	0.003848	0.003885
53	0.003483	0.003753	0.003865
54	0.003367	0.003721	0.003819
55	0.003277	0.003679	0.003785
56	0.003164	0.003656	0.003741
57	0.003196	0.003631	0.003738
58	0.003169	0.003662	0.003730
59	0.003198	0.003621	0.003733
60	0.003252	0.003626	0.003776
61	0.003301	0.003630	0.003755
62	0.003361	0.003695	0.003766
63	0.003487	0.003735	0.003739
64	0.003519	0.003736	0.003789
65	0.003578	0.003798	0.003807
66	0.003683	0.003792	0.003826
67	0.003753	0.003833	0.003866
68	0.003785	0.003852	0.003848
69	0.003900	0.003841	0.003867
70	0.003928	0.003900	0.003906
71	0.003970	0.003917	0.003896
72	0.004039	0.003933	0.003958
73	0.004061	0.003946	0.003936
74	0.004065	0.003983	0.003929
75	0.004050	0.003963	0.003951
76	0.004045	0.003975	0.003953
77	0.004005	0.003959	0.003940
78	0.004019	0.003971	0.003921
79	0.004005	0.003925	0.003959
80	0.003976	0.003920	0.003920
81	0.003939	0.003921	0.003920
82	0.003913	0.003910	0.003891
83	0.003938	0.003902	0.003940
84	0.003873	0.003869	0.003893
85	0.003890	0.003895	0.003892
86	0.003879	0.003877	0.003864
87	0.003895	0.003886	0.003897
88	0.003877	0.003892	0.003897
89	0.003883	0.003879	0.003883
90	0.003893	0.003851	0.003879
91	0.003890	0.003904	0.003917
92	0.003883	0.003880	0.003918
93	0.003904	0.003892	0.003907
94	0.003907	0.003894	0.003904
95	0.003889	0.003877	0.003897
96	0.003889	0.003920	0.003908
97	0.003909	0.003895	0.003904
98	0.003866	0.003919	0.003915
99	0.003910	0.003929	0.003892
100	0.003913	0.003895	0.003878
101	0.003879	0.003912	0.003877
102	0.003925	0.003889	0.003855
103	0.003912	0.003899	0.003894
104	0.003901	0.003911	0.003904
105	0.003911	0.003900	0.003884
106	0.003900	0.003901	0.003935
107	0.003891	0.003899	0.003929
108	0.003887	0.003887	0.003893
109	0.003918	0.003879	0.003908
110	0.003881	0.003881	0.003900
111	0.003897	0.003876	0.003873
112	0.003894	0.003920	0.003864
113	0.003933	0.003895	0.003889
114	0.003919	0.003951	0.003864
115	0.003901	0.003903	0.003907
116	0.003904	0.003921	0.003950
117	0.003915	0.003937	0.003931
118	0.003884	0.003900	0.003877
119	0.003882	0.003907	0.003910
120	0.003921	0.003929	0.003924
121	0.003894	0.003886	0.003895
122	0.003911	0.003867	0.003927
123	0.003906	0.003907	0.003914
124	0.003908	0.003885	0.003933
125	0.003879	0.003894	0.003934
126	0.003898	0.003902	0.003889
127	0.003934	0.003867	0.003915

Table 3 continues in the next page.

y	$P(S_N[y] = f_y)$	$P(S_N^2[y] = f_y)$	$P(S_N^3[y] = f_y)$
128	0.003881	0.003906	0.003913
129	0.003851	0.003897	0.003887
130	0.003863	0.003858	0.003905
131	0.003862	0.003873	0.003901
132	0.003879	0.003906	0.003927
133	0.003875	0.003864	0.003897
134	0.003895	0.003891	0.003905
135	0.003876	0.003849	0.003893
136	0.003904	0.003888	0.003883
137	0.003869	0.003861	0.003891
138	0.003845	0.003918	0.003843
139	0.003857	0.003897	0.003932
140	0.003890	0.003896	0.003897
141	0.003884	0.003857	0.003938
142	0.003892	0.003887	0.003870
143	0.003909	0.003918	0.003895
144	0.003927	0.003885	0.003901
145	0.003885	0.003848	0.003921
146	0.003893	0.003874	0.003928
147	0.003858	0.003887	0.003923
148	0.003900	0.003886	0.003917
149	0.003861	0.003892	0.003877
150	0.003901	0.003927	0.003903
151	0.003906	0.003895	0.003884
152	0.003894	0.003862	0.003895
153	0.003887	0.003937	0.003897
154	0.003907	0.003895	0.003902
155	0.003894	0.003898	0.003888
156	0.003908	0.003901	0.003928
157	0.003903	0.003863	0.003895
158	0.003931	0.003912	0.003910
159	0.003925	0.003890	0.003883
160	0.003938	0.003898	0.003894
161	0.003892	0.003901	0.003922
162	0.003911	0.003899	0.003910
163	0.003893	0.003913	0.003933
164	0.003873	0.003901	0.003883
165	0.003892	0.003900	0.003889
166	0.003904	0.003895	0.003883
167	0.003906	0.003875	0.003893
168	0.003886	0.003909	0.003884
169	0.003904	0.003903	0.003915
170	0.003903	0.003901	0.003890
171	0.003888	0.003927	0.003880
172	0.003930	0.003899	0.003906
173	0.003919	0.003955	0.003906
174	0.003898	0.003930	0.003919
175	0.003927	0.003919	0.003941
176	0.003899	0.003889	0.003906
177	0.003912	0.003906	0.003899
178	0.003893	0.003917	0.003902
179	0.003903	0.003891	0.003915
180	0.003920	0.003908	0.003888
181	0.003914	0.003899	0.003934
182	0.003918	0.003875	0.003938
183	0.003902	0.003932	0.003872
184	0.003916	0.003902	0.003913
185	0.003901	0.003914	0.003933
186	0.003887	0.003926	0.003905
187	0.003911	0.003895	0.003904
188	0.003925	0.003881	0.003878
189	0.003903	0.003888	0.003880
190	0.003914	0.003923	0.003898
191	0.003916	0.003897	0.003876
192	0.003910	0.003862	0.003922
193	0.003930	0.003895	0.003924
194	0.003852	0.003905	0.003911
195	0.003906	0.003873	0.003909
196	0.003914	0.003878	0.003893
197	0.003947	0.003932	0.003909
198	0.003955	0.003908	0.003889
199	0.003919	0.003925	0.003912
200	0.003897	0.003908	0.003897
201	0.003908	0.003918	0.003928
202	0.003935	0.003918	0.003904
203	0.003919	0.003916	0.003891
204	0.003951	0.003881	0.003944
205	0.003882	0.003916	0.003915
206	0.003887	0.003849	0.003969
207	0.003904	0.003919	0.003881
208	0.003899	0.003885	0.003910
209	0.003896	0.003885	0.003867
210	0.003914	0.003888	0.003915
211	0.003911	0.003916	0.003931
212	0.003907	0.003907	0.003899
213	0.003884	0.003895	0.003898
214	0.003884	0.003897	0.003898
215	0.003896	0.003896	0.003941
216	0.003918	0.003916	0.003910
217	0.003931	0.003876	0.003926
218	0.003908	0.003880	0.003927
219	0.003906	0.003901	0.003894
220	0.003937	0.003947	0.003923
221	0.003904	0.003915	0.003922
222	0.003886	0.003926	0.003906
223	0.003893	0.003890	0.003884
224	0.003928	0.003881	0.003939
225	0.003931	0.003929	0.003892
226	0.003882	0.003876	0.003913
227	0.003902	0.003910	0.003875
228	0.003922	0.003932	0.003883
229	0.003933	0.003920	0.003881
230	0.003884	0.003933	0.003936
231	0.003932	0.003910	0.003894
232	0.003897	0.003858	0.003897
233	0.003911	0.003871	0.003898
234	0.003936	0.003921	0.003877
235	0.003930	0.003908	0.003919
236	0.003914	0.003924	0.003902
237	0.003891	0.003891	0.003897
238	0.003908	0.003868	0.003891
239	0.003930	0.003911	0.003915
240	0.003953	0.003914	0.003863
241	0.003901	0.003926	0.003918
242	0.003875	0.003926	0.003880
243	0.003934	0.003910	0.003898
244	0.003888	0.003927	0.003937
245	0.003914	0.003887	0.003895
246	0.003896	0.003892	0.003902
247	0.003913	0.003886	0.003915
248	0.003878	0.003937	0.003907
249	0.003867	0.003911	0.003901
250	0.003904	0.003888	0.003904
251	0.003885	0.003921	0.003919
252	0.003932	0.003882	0.003884
253	0.003901	0.003886	0.003923
254	0.003913	0.003907	0.003917
255	0.003903	0.003885	0.003929

Table 3: Experimental Results: $P(S_N[y] = f_y)$, $P(S_N[S_N[y]] = f_y)$ and $P(S_N[S_N[S_N[y]]] = f_y)$ versus y for $0 \leq y \leq 255$.