SOME COMBINATORIAL ASPECTS OF THE KSA\textsubscript{m}-LIKE
ALGORITHMS SUITABLE FOR RC4 STREAM CIPHER

BOGDAN CRAINICU, FLORIAN MIRCEA BOIAN

Abstract. RC4 remains one of the most widely used stream cipher. In order to face the main critical weaknesses, a number of proposals for modifying RC4 algorithm have been advanced. In this paper we analyze some combinatorial aspects regarding the randomness of a variant of the Key-Scheduling Algorithm (KSA), called KSAm, proposed by Crainicu and Boian in [2] as a better protection against Initialization Vectors (IVs) weakness of Wired Equivalent Privacy (WEP) cryptosystems. Based on a model presented by Mironov in [19], we calculate the sign of the entries’ permutation of the internal state table \( S \) after KSAm, which provides a negligible advantage of guessing a particular bit. Then, we analyze the probability of the event where a particular initial value follows a linear forward movement through the vector \( S \), with possible undesirable consequences in predicting the value during that movement.

1. Introduction

RC4 is a stream cipher which was designed by Ron Rivest in 1987 for RSA Security. RC4 was kept as a trade secret until an alleged copy of it was anonymously posted to the Cypherpunks mailing list in 1994.

Because of its simplicity and speed, RC4 is one of the most widely used stream cipher; for example, it is used in the SSL/TLS (Secure Socket Layer/Transport Layer Security) standards, WEP (Wired Equivalent Privacy), WPA (Wi-Fi Protected Access), and it can be also found in email encryption products.

There were discovered many significant weaknesses of RC4 and RC4-based WEP implementations: weak IVs/keys [1, 4, 8, 9, 10, 13, 14, 21, 22, 26, 28], invariance weakness [4], bias in the second output [16], related key attack [4, 27].
7], state recovery attack [11, 20, 23, 27], distinguishing attack [3, 5, 16-19, 24, 25], biased distribution of RC4 initial permutation [24, 28].

In order to address its critical weaknesses, a number of proposals for modifying RC4 algorithm have been advanced: Paul and Preneel present in [25] a new pseudorandom bit generator called RC4A, Zoltak proposes in [29] the VMPS stream cipher, and Gong, Gupta, Hell and Nawaz also propose in [6] a new 32/64-bit RC4-like keystream generator.

Crainicu and Boian proposed in [2] a modified version of KSA, called KSAm, whose primary goal was to face the weakness exhibited by Fluhrer, Mantin and Shamir (FMS) in [4], where certain IVs leaks information about the secret key in WEP mode of operation. The authors demonstrate that the attacker has no possibilities to manipulate KSAm permutation in order to reach the FMS resolved condition. Based on the Roos’ experimental observation [26], there is a weaker probabilistic correlation between the first three words of the secret key and the first three entries of the state table after KSAm, which causes a negligible bias of the first word of the RC4KSAm (RC4 with KSAm as Key-Scheduling Algorithm) output stream towards the sum of the first three words of the secret key. The effect of this negligible bias can be easily avoided by discarding only the first word from the RC4KSAm output stream.

In this paper, we examine two combinatorial properties of the KSAm in its normal mode of operation, and not from the perspective of a particular implementation. Firstly, based on a model of a shuffling technique presented by Mironov in [19], where an idealized RC4 stream cipher is involved, we compute the sign of the permutation of $S$ after KSAm, whose values help to predict a bit $b$ with a probability of 0.91% over a random guess. This advantage is too small to be feasible in an attack. We also analyze the state table entries during the KSAm steps, with special focus on calculating the probability of a linear advance movement of an initial value from a particular state table entry during KSAm. The results prove that it is very unlikely to find a location in $S$ during such movement where that value may be predicted with a probability significantly greater than 1/N.

2. KSAm

Crainicu and Boian suggest in [2] a modified version of the original KSA, called KSAm (Fig. 1), for addressing the FMS weakness of WEP-like cryptosystems, mainly when IV precedes the secret key.

The KSAm takes the secret key and initializes a vector of indices $u_0, u_1, \ldots, u_{N-1}$; the values of indices $u_i$ are not necessarily unique within the vector of indices, and they are kept secret. Then, it swaps the two values of $S$ pointed to
KSA \((K, S)\)

Initialization:
for \(i = 0\) to \(N - 1\)
\(S[i] = i;\)

Scrambling:
for \(i = 0\) to \(N - 1\)
\(j = (j + S[i]) \mod N;\)

KSAm \((K, S)\)

Initialization:
for \(i = 0\) to \(N - 1\)
\(S[i] = i;\)

Scrambling 1:
for \(i = 0\) to \(N - 1\)
\(u_i = (S[i] + K[i \mod \ell]) \mod N;\)

Scrambling 2:
for \(i = 0\) to \(N - 1\)
\(j = (j + S[i]) \mod N;\)

\[\text{by } i \text{ and } u_i, \text{ so that the Scrambling 1 stage of KSAm ends with a secret state, which is different from the identity permutation with a very high probability. The rest of operations (Scrambling 2) remain the same as in the original KSA: it applies the scrambling rounds } N = 2^n \text{ times, stepping } i \text{ across } S, \text{ updating } j \text{ by adding the previous value of } j, S[i] \text{ and the next word of the key.}\]

In fact, KSAm comprises a family of key scheduling algorithms, where Scrambling 1 sequence tries to follow the Knuth’s observation \([12]\): instead of swapping \(S[i]\) with a random entry, it must be swapped with an entry randomly chosen from \(S[i]\) to \(S[N - 1]\) (the implementation of this concept remains still problematic due to the randomness of the secret key \(K\)).

At a glance, the first observation is that there are now two different scrambling processes, both of them based on the same secret key. Even if the computation/running time of KSAm is almost twice as long as that of KSA, the additional time is insignificant (the software implementation remains very fast).

The security of KSAm comes also from its huge internal state. The internal state of \(RC4_{\text{KSA}}\) is approximately 1700 bits for 8-bits words. Instead, KSAm provides a much larger size and, as result, it is much harder to reconstruct its internal state. Crainicu and Boian present in \([2]\) the formula for calculating \(L_{\text{RC4–KSAm}}\), which represents the size of the \(RC4_{\text{KSAm}}\’\) internal state (the values of indices \(u_i\) are not necessarily unique; therefore, the number of all possibilities of distributing \(2^n\) elements into \(2^n\) cells where repetitions are allowed is \((2^n)^{2^n}\) \([2]\):

\[
\text{Figure 1. KSA vs KSAm [2]}\]
$$L_{RCA-KSAM} = \log_2(2^n! \times (2^n)^2 \times (2^n)^2) = [\log_2(2^n!) + (n \times 2^n) + 2n]$$
$$L_{RCA-KSAM,n=8} \approx 3748 \text{ bits}$$

Beside KSA, KSAm needs only additional 256 bytes of memory for the indices $$u_i$$ (for $$N = 8$$), which represents a negligible amount of supplementary memory.

3. **On Randomness of KSAm**

Two independent scrambling processes are involved with KSAm; therefore, after running consecutively both of them, each element of the state table will be swapped at least twice (possibly with itself).

Based on a series of significant studies on the original KSA [5, 16, 17, 19, 20, 23, 24], two of the most important approaches for analyzing KSAm are to deduce the probability of a linear advance movement of a value $$b$$ which each step along the locations of vector $$S$$, and also to calculate the probability of a value $$b$$ to end up in any location $$a$$ (including the probability of identity permutation).

3.1. **The sign of the permutation $$S$$ after KSAm.** Mironov calculates in [19] the limiting distribution for the two possible values (+1 and −1) of the sign of $$S$$ after KSA:

$$P(sign(S) = (-1)^N) = \frac{1}{2} \left(1 + e^{-2}\right)$$
$$P(sign(S) = (-1)^{N-1}) = \frac{1}{2} \left(1 - e^{-2}\right)$$

Therefore, Mironov demonstrates that it is possible to predict the sign of the permutation $$S$$ after KSA with probability $$1 - \frac{1}{2} \cdot e^{-2}$$, and he shows that this value of about 6.7% over a random guess becomes also the advantage of guessing the bit $$b$$ correctly.

Next, we compute the sign of the permutation $$S$$ after KSAm. Two scrambling sequences are involved in KSAm, and therefore we have $$2N$$ rounds. At each round and regardless of what scrambling sequence is about, the swap process changes the sign of the permutation only if $$i \neq j$$. This happens with probability $$\left(1 - \frac{1}{N}\right)$$. If $$i = j$$, with probability $$\frac{1}{N}$$, the values of $$S$$ pointed by $$i$$ and $$j$$ remain unchanged. The probability that $$i \neq j$$ during all $$2N$$ rounds of KSAm is $$\left(1 - \frac{1}{N}\right)^{2N}$$, and the probability that $$i = j$$ during all $$2N$$ rounds of KSAm is $$\left(\frac{1}{N}\right)^{2N}$$. The sign of the identity permutation is +1.

Based on these observations, the probability that the sign of $$S$$ is changed an even number of times and at the end of KSAm is negative, is:
The probability that the sign of \( S \) is changed an odd number of times and at the end of KSAm is negative, is:

\[
P(\text{sign}(S) = (-1)^{2N}) = 1 - P(\text{sign}(S) = (-1)^{2N-2}) = 1 - \frac{1}{2} \cdot (1 - e^{-\frac{4}{2}}) = 1 - \frac{1}{2} \cdot (1 - e^{-4})
\]

After KSAm, the sign of \( S \) can be predicted with an very small advantage of \( e^{-4} \approx 0.0091 \) over the random guess, which means approximately 0.91%.

Mironov obtains in [19] approximately the same result by running consecutively two times the original KSA.

Even if we found a small bias towards the sign of the permutation after KSAm, the value \( e^{-4} \) is totally useless for attacking RC4 based on KSAm. As further precaution, discarding only the first three words of the output of RC4 ensures the security of the algorithm.

3.2. Probability of a linear advance movement of an initial value from a particular state table entry during KSAm. We define the minimum probability of a given initial entry \( S_0[a] = a \) as the probability that this entry remains unchanged during each step of one of the two scrambling processes of KSAm. Thus, the minimum probability of the identity permutation is defined as the multiplication of all minimum probabilities corresponding to each initial entry \( S_0[a] = a \).

We refine the Theorem 1 from [2]:

\[
P(\text{sign}(S) = (-1)^{2N-1}) = 1 - P(\text{sign}(S) = (-1)^{2N}) = 1 - \frac{1}{2} \cdot (1 - e^{-4})
\]
Theorem 1. The minimum probability of a given initial entry $S_0[a] = a$ after $N$ steps is:

$$P(S_N[a] = a) = \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^{N-1}$$

Proof. At some point, the index $i$ touches the value $a$. In this round, with probability $1/N$, $i = j = a$, and therefore $S[a]$ will be swapped with itself. For the rest of the $(N-1)$ rounds we have $i \neq a$, and $j \neq a$ with probability $(1 - \frac{1}{N})$.

The minimum probability of the identity permutation, which means that all $N$ entries of table $S$ remain unchanged after completion of one of the two scrambling process is:

$$P(S_N = \text{identity permutation}) = \left[\frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^{N-1}\right]^N$$

Result 1. [19]: The probability of the value $b$ to end up in location $a$ in $S$ at the end of the KSA round (the distribution of KSA outputs) is:

$$P[S_N[a] = b] = \begin{cases} 
\frac{1}{N} \left[\left(1 - \frac{1}{N}\right)^{N-a-1} + \left(1 - \frac{1}{N}\right)^b\right] & \text{if } a < b \\
\frac{1}{N} \left[\left(1 - \frac{1}{N}\right)^b + \left(1 - \frac{1}{N}\right)^b \left(1 - \frac{1}{N}\right)^{N-a-1}\right] & \text{if } a \geq b 
\end{cases}$$

For example, the minimum probability for event $S_N[0] = 0$ is $\frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^{N-1}$, and according to Result 1, the probability for event $S_N[0] = 0$ is $\frac{1}{N}$.

Theorem 2. Assuming that both indices $u_i$ and $j$ take values independently and uniformly at random at each round of the two scrambling processes of KSAm, the probability of the value $b$ to end up in location $a + 1$ in $S$ in the round $a$ of Scrambling 1 or Scrambling 2, where $S_0[0] = b$, is:

$$P(S_a[(a+1) \mod N] = b) = \frac{1}{N}, a \in [1, N] \text{ and } b \in [0, N-1]$$

Proof. At each round, taking into account that the value of $i$ is known, we can analyze the probability of the event $S_a[(a+1) \mod N] = b$, which depends on the value taken by $u_i$ or $j$. The value $b$, during the $a$ rounds, must not remain behind a location pointed by $i(i \in [0, (a+1) \mod N])$ nor to reach a position after the $[(a+1) \mod N]^{th}$ location in $S$, because, in these situations, there are no possibilities to manipulate the value $b$ so that it ends
up in location \([(a + 1) \mod N]\). Following the steps of either *Scrambling 1* or *Scrambling 2*, we have:

\[
S_1[2] = b \text{ if } j_1 = 2 \text{ (the initial condition is } S_0[0] = b) \Rightarrow P(S_1[2] = b) = P(j_1 = 2) = \frac{1}{N}; \\
S_2[3] = b \text{ if } < j_1 = 1 \text{ and } j_2 = 3 > \text{ or } < j_1 = 3 \text{ and } j_2 \neq 3 > \Rightarrow P(S_2[3] = b) = P(j_1 = 1) \cdot P(j_2 = 3) + P(j_1 = 3) \cdot P(j_2 \neq 3) = \\
\frac{1}{N^2} + \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right) = \frac{1}{N}; \\
S_3[4] = b \text{ if } < j_1 = 1 \text{ and } j_2 = 2 \text{ and } j_3 = 4 > \text{ or } < j_1 = 1 \text{ and } j_2 = 4 \text{ and } j_3 \neq 4 > \Rightarrow P(S_3[4]) = P(j_1 = 1) \cdot P(j_2 = 2) \cdot P(j_3 = 4) + P(j_1 = 1) \cdot P(j_2 = 4) \cdot P(j_3 \neq 4) + P(j_1 = 2) \cdot P(j_2 \neq 2) \cdot P(j_3 = 4) + P(j_1 = 4) \cdot P(j_2 \neq 4) \cdot P(j_3 \neq 4) = \\
\frac{1}{N^3} + \frac{2}{N^2} \cdot \left(1 - \frac{1}{N}\right) + \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^2 = \frac{1}{N}; \\
\ldots \\
P(S_4[5] = b) = \frac{1}{N^4} + \frac{3}{N^3} \cdot \left(1 - \frac{1}{N}\right) + \frac{3}{N^2} \cdot \left(1 - \frac{1}{N}\right)^2 + \\
+ \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^3 = \frac{1}{N} \\
P(S_5[6] = b) = \frac{1}{N^5} + \frac{4}{N^4} \cdot \left(1 - \frac{1}{N}\right) + \frac{6}{N^3} \cdot \left(1 - \frac{1}{N}\right)^2 + \\
+ \frac{4}{N^2} \cdot \left(1 - \frac{1}{N}\right)^3 + \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^4 = \frac{1}{N}; \\
\ldots
\[
P(S_a[a+1] = b) = \frac{1}{N} \cdot \left(1 - \frac{1}{N}\right)^{a-1} + \frac{(a-1)}{1! \cdot N^2} \cdot \left(1 - \frac{1}{N}\right)^{a-2} + \frac{(a-1) \cdot (a-2)}{2! \cdot N^3} \cdot \left(1 - \frac{1}{N}\right)^{a-3} + \frac{(a-1) \cdot (a-2) \cdot (a-3)}{3! \cdot N^4} \cdot \left(1 - \frac{1}{N}\right)^{a-4} + \frac{(a-1) \cdot (a-2) \cdot (a-3) \cdot (a-4)}{4! \cdot N^5} \cdot \left(1 - \frac{1}{N}\right)^{a-5} + \ldots + \frac{1}{N^a} = \frac{1}{N}
\]

For Scrambling 1, \( b = 0 \), and for Scrambling 2, \( b \in [0, 255] \). Applying the Result 1, where the initial state is the identity permutation, for \( b = 0 \) and \( a = 1 \), we obtain \( P(S_N[1] = 0) = \frac{1}{N} \), and applying the Theorem 2 for \( a = N \), we have the same result: \( P(S_N[(N+1)\mod N] = 0) = \frac{1}{N} \).

Theorem 2 can be adapted for all entries of the initial permutation \( S_0 \). The significance of Theorem 2 lies in the fact that the event that a particular value \( b \) follows a linear path through the vector \( S \), and consequently ends up in an expected location after Scrambling 1, has a probability around \( \frac{1}{N} \). This result has also to be considered in the context of starting Scrambling 2 with a state table \( S \) which is different from the identity permutation with high probability.

4. Conclusions

We investigated KSAm [2] from the point of view of shuffling algorithm presented by Mironov in [19]. We calculated as well the probability of the sign of the permutation \( S \), but after KSAm, finding a value which may help in predicting that sign with an advantage of 0.91% over a random guess. Mironov obtains in [19] about the same result by running consecutively two times the original KSA, but KSAm benefits from a much larger size of the internal state and the running of two different scrambling processes. The mentioned advantage of 0.91% over a random guess, biased towards the sign of the permutation after KSAm, is though too small, so that an attack against RC4\textsubscript{KSAm} could not rely on it. As precaution, an additional measure for thwarting against this weakness consists in discarding only the first three words of the output of RC4\textsubscript{KSAm}.

The second part of our analysis is focused on calculating the probability of a particular event, namely, a linear advance movement of the state table entry \( S_0[0] = b \) during KSAm rounds. The value obtained is about \( \frac{1}{N} \), which demonstrates that such event happens randomly. The result can be extended to the others entries of the initial state table \( S \).
REFERENCES


[26] A. Roos, Class of weak keys in the RC4 stream cipher, Two posts in sci.crypt, message-id 43ulhe81j30hermes.is.co.za and 44ebg@ffh@hermes.is.co.za, 1995.

"Petru Maior" University of Tîrgu-Mureş
E-mail address: cbogdan@upm.ro