8th International ISC Conference
on Information Security and Cryptology
(ISCISC'11)

Ferdowsi University of Mashhad-September 14-15, 2011

On the Period of GSM's A5/1 Stream Cipher and Its Internal State Transition Structure

Vahid Amin Ghaffari

ICT
Malek Ashtar University of Technology
Tehran, Iran
vahidaming@yahoo.com

Ali Vardasbi

Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
vardasbi@ee.sharif.edu

***Abstract-* The A5/1 algorithm is one of the most applicable and famous stream cipher algorithms. However, the period of A5/1 keystream sequence and the exact structure of its internal state transition were not investigated thoroughly, until now. This paper deals with the transition of the internal state of A5/1 algorithm and tries to present a model for such a transition. Additionally, the internal states are divided into two groups, initially periodic and ultimately periodic. The presented model is verified using a variety of simulations and it is consistent with theoretical results as well.**

Keywords- A5/1 algorithm, internal state transition, ultimately periodic

I. INTRODUCTION

The family of A5 stream ciphers is one of the tools used in GSM protocol to make the phone calls secure. Nowadays there are more than one billion people around the world, using A5/1 algorithm in their cell phones [1]. Starting from 1997, this algorithm was analyzed by many researchers, among which were renown cryptanalysts like Shamir, Biham, Biryukov and Golić [1,2,3].

The problem of specifying the period of a stream cipher's keystream sequence and its internal state is crucial in evaluating the security of that stream cipher. Although some references mentioned the period of A5/1 keystream sequence to be approximately 2^{23}, there is no published paper about the exact structure of A5/1 internal state transition.

The A5/1 stream cipher consists of three LFSRs with 19, 22 and 23 bits length. In each clocking of the algorithm, a control bit is taken from a fixed position of each LFSR. Then the LFSRs are clocked, due to a majority logic obtained from the three control bits[1].

Studying the state transition of this algorithm is useful in some cryptanalysis methods such as rainbow attacks, in which knowing the probability of a collision in the produced table, could be of great importance [1]. If, for example, there was a feasible way for distinguishing the states which can be back clocked to more than 100 previous states, from the states lacking this property; then, choosing the starting points in the rainbow tables could be done more efficiently. This could make better the results of time-memory tradeoff attacks in [2], where the states which can be clocked backwards a minimum of 100 clocks and the ones which cover more states after 328 back-clocking are preferred.

The rest of this paper is organized as follows. A review of the previous analysis of A5/1 algorithm is given in section two. Section three is devoted to the investigation of state transition in A5/1 algorithm, while section four contains the related theoretical view. Section five will conclude the paper.

II. REVIEW OF A5/1 ALGORITHM CRYPTANALYSIS

The first official cryptanalysis on A5/1 stream cipher was presented in 1997 by Golić, who proposed two known plain text attacks on the cipher: guess and determine and time-memory tradeoff attack [3]. However, neither of those attacks was practical.

Biryukov, Shamir and Wagner presented another time-memory tradeoff attack on A5/1 in 2000 [2]. Their cryptanalysis requires a large amount of known plain text as well as a high pre-computation time; hence, is not practical. At the same time, Biham and Dunkelman presented a different cryptanalysis in which some improved tables were exploited in order to perform a guess and determine attack on A5/1 [4]. This attack suffers from requiring too many known plain texts, too.

Later on, three cryptanalysis were proposed on A5/1, in the form of correlation attacks; but they required too many known plain or cipher texts [5, 6], or they assumed ideal conditions for the attack [7], which made them impractical like the previous ones.

Barkan, Biham and Keller proposed a time-memory-data tradeoff attack on A5/1 in 2007 [1]. Although this attack contains a high complexity pre-computation phase, it seems to be more practical than the previous attacks.

III. STATE TRANSITION IN A5/1 ALGORITHM

If A5/1's registers were not clocked with respect to a majority function; i.e. all the three LFSRs were clocked in all

[1] The clocking of the algorithm should not be confused with each LFSR's clocking.

the algorithm clocks, due to the LFSRs' primitive characteristic function and their relatively prime size, the period of the algorithm's generated keystream would be $\approx 2^{64}$. However, the majority function makes it hard to comment about the period of the keystream sequence. In [8], the period of an algorithm "like A5/1" was observed to be near $\frac{4}{3}(2^{23}-1)$. This observation was later referenced by Golić in [3] for A5/1 algorithm.

Our investigation, however, showed that, with a high probability, a randomly selected initial state will never be repeated; suggesting the keystream sequence is ultimately periodic. We tested a set of 10000 randomly selected initial states and in neither of them the first 64 keystream bits were repeated.

A. *Loop states and ultimately periodic states*

After a large amount of experiments and simulations on the internal state transition of the A5/1 algorithm, the states were observed to enter a loop[1] after an average number of $2^{26.17}$ algorithm clocks. These simulations showed that a big proportion of all the internal states will never be repeated. In other words, the loop states (i.e. the states which are repeated during the run of the algorithm) constitute only a small part of the internal states.

Therefore, the space of the algorithm's internal states could be divided into some independent pages, each page containing one loop, in which there are some branches entering the loop. A scheme of the internal states in a page is depicted in figure 1.

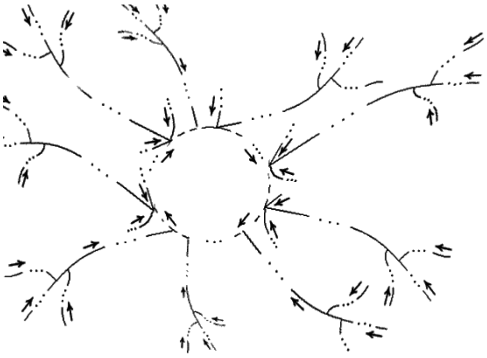

Figure 1. A scheme of the internal states in a page

TABLE I. THE MINIMUM AND MAXIMUM OF THE DISTANCE OF EACH STATE TO A LOOP AND THE PERIOD OF THE LOOP

	distance to loop	Loop period
Sample with maximum period of the loop	$2^{26.19}$	$2^{28.41}$
Sample with minimum period of the loop	$2^{27.32}$	$2^{23.41}$
Sample with maximum distance to the loop	$2^{28.08}$	$2^{23.41}$
Sample with minimum distance to the loop	$2^{17.27}$	$2^{24.41}$

All the states in each page will ultimately go into a loop. Throughout this paper, the number of clocks after which a state meets the loop is called the distance of that state to its loop.

In one of our simulations, 100 initial states were selected randomly and the distance of each state to a loop as well as the period of the loop was measured. The average distance of a randomly selected state to its loop was $2^{26.17}$, while the average period of each loop was $2^{25.42}$. Table I shows the min-imum and maximum of the measured values.

Table II contains the ratio of the number of states with one, two, three and four possible predecessors to the total states in a loop. It is interesting to note that despite the different period of loops, this ratio remains to be approximately constant for all the loops.

TABLE II. THE RATIO OF THE STATES WITH ONE, TWO, THREE AND FOUR POSSIBLE PREDECESSORS TO THE TOTAL STATES IN A LOOP

One state	Two states	Three states	Four states
0.65	0.125	0.156	0.062

Our results are consistent with the following proposition about the number of possible predecessors [3]. Table III shows the theoretical ratio of the number of states with one, two, three and four possible predecessors to the 62.5% states which have at least one possible state.

TABLE III. THE THEORETICAL RATIO OF THE STATES WITH ONE, TWO, THREE AND FOUR POSSIBLE PREDECESSORS TO THE STATES WITH AT LEAST ONE PREDECESSOR

One state	Two states	Three states	Four states
0.65	0.15	0.15	0.05

Proposition 1: If an internal state $S(t)$ is randomly chosen according to uniform distribution, then the number of solutions for the predecessor $S(t-1)$ is a nonnegative integer random variable Z with the probability distribution

$$\Pr\{Z=0\}=\frac{3}{8},\ \Pr\{Z=1\}=\frac{13}{32},$$
$$\Pr\{Z=2\}=\Pr\{Z=3\}=\frac{3}{32},\ \Pr=\{Z=4\}=\frac{1}{32} \tag{1}$$

It is notable that 37.5% of all the states have no possible predecessors. In other words, 37.5% of all the states could only be a starting point and they never are produced. This was noted in [2] and the paper was based on the elimination of the states with no possible states after 100 backward clocking and exploiting the states which cover more possible states when they are clocked back. Obviously, the states in a loop are more suitable for being used in time-memory-data tradeoff tables, due to their better coverage of the states when they are clocked back.

Several statistical tests such as ordinary test, NIST tests and overall test were applied on the states both in and out of the loop in order to find a way for distinguishing the loop states

[1] Some references call it a "cycle"

from the other ones. Unfortunately, these tests had no significance in distinguishing between the states in the loops and the other states.

B. *Number of possible predecessors*

Proposition 1 only deals with one step back-clocking. Following a similar approach as the one in [3], one can propose the distribution of number of possible predecessors for more than only one step of back-clocking. In other words, one can consider all the $2^{3(d+1)}$ different values for $d+1$ bits of each LFSR in A5/1 which is shown in figure 2 as the shaded area.

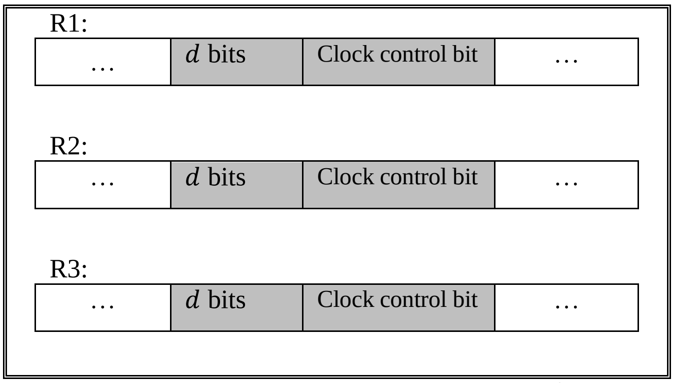

Figure 2. Changing bits in three LFSRs to obtain number of possible predecessors in d clocks back

Then, each one of $2^{3(d+1)}$ states should be back-clocked for d clocks and the number of possible predecessors should be monitored. At the end, the ratio of all the states with k ($k = 0, \dots, 4^d$) possible predecessors to all the $2^{3(d+1)}$ states is calculated. Table IV shows the percentage of states with k ($k = 0,1,2$) predecessors for d ($d = 1, \dots, 6$) clocks back.

Table IV shows the sum of ratios for $k = 0$ and $k = 1$ is almost constant for different "d"s, that is 80%. Furthermore, an approximate of 94% of all states have less than four possible predecessors for different numbers of clocks back.

TABLE IV. THE THEORETICAL PERCENTAGE OF STATES WITH k ($k = 0,1,2$) PREDECESSORS FOR d ($d = 1, \dots, 6$) CLOCKS BACK

	$k = 0$	$k = 1$	$(k = 0)$ & $(k = 1)$ & $(k = 2)$ & $(k = 3)$
$d = 1$	37.5	40.6	96.9
$d = 2$	42.2	37.9	95.3
$d = 3$	43.9	35.8	94.9
$d = 4$	45.3	34.2	94.7
$d = 5$	46.6	32.8	94.4
$d = 6$	47.9	31.5	94.1

C. *Number of different pages*

Our simulations led us to the conclusion that the average number of each page's states is approximately $2^{51.6}$ and consequently, there are about $2^{12.4}$ pages. Here is a brief explanation on the number of pages.

The first step is to determine the number of each loop's branches. The states in a loop can be categorized into four groups, based on the number of their possible predecessors. Each state in group g ($g = 1, \dots, 4$), introduces $g - 1$ branches to the loop. Therefore, one can estimate the number of branches in a loop as follows:

$$B = L \cdot \sum_{g=1}^{4} \{\Pr(Z = g) \cdot (g - 1)\} \qquad (2)$$

where B is the average number of branches in one loop and L is the average number of states in each loop.

Using the values in table III, and $L \approx 2^{25.42}$ the average number of branches in a loop would be:

$$B \approx 2^{24.68} \qquad (3)$$

Now, only the average number of states in each branch (which we denote it by n) needs to be obtained. Based on our simulations, the average distance of a state to its loop is $2^{26.17}$; we call this quantity D. In order to estimate n, the number of states in each branch, it should be noticed that an approximate of 40% of the states have no possible predecessors, hence being on the starting points of the branches. Assuming this is the case for the points in each branch (i.e. around 40% of all the points of each branch are the starting points), the distance of each point to the loop would be through the remaining 60% of the branch points. This observation suggests that D constitutes about 60% of n, leading to the following conclusion:

$$n \approx \frac{D}{0.6} \approx 2^{26.91} \qquad (4)$$

Finally, each page contains an average of N states, where N could be estimated by:

$$N = n \cdot B + L \approx 2^{51.6} \qquad (5)$$

and there are approximately $\frac{2^{64}}{2^{51.6}} = 2^{12.4}$ different pages. It has to be noted that this approach gave us a rough estimation and more works may be required in order to better these estimations.

IV. THEORETICAL VIEW OF THE RESULT

In explaining this situation for the internal states of the A5/1 algorithm, two points should be noted. First, there are a finite number of internal states and this means, at some point, the internal state will be repeated; i.e. the internal states sequence is ultimately periodic. Second, 37.5% of the states have no possible predecessors, meaning that these states could only be the starting points and they will never be generated in the clocking process (an example of this status is shown in figure 3). Therefore, the scheme in figure 1 is theoretically explainable too. This behavior may also be observed in nonlinear shift registers.

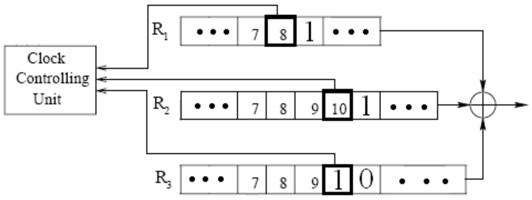

Figure 3. An example of the states have no possible predecessors

V. Conclusion

In this paper some discussions on the state transition of A5/1 algorithm was presented. One of the important applications of this discussion may be in time-memory-data tradeoff attacks (especially rainbow attacks), where the states with possible previous 100 states and the ones which cover more states after 328 backwards clocking are preferred.

In investigating the properties of the period of keystream sequence (and so the internal states), some simulations were performed which in one of them a set of 10000 randomly selected samples was tested. It was observed that none of these samples were repeated in the process of clocking. A wide range of experiments has led us to the conclusion that after an average of approximately $2^{26.17}$ clocks, the algorithm's internal states will enter some loop and only the loop states are repeated during the process of clocking. Based on this observation, the basic scheme of the structure of the internal states transition was presented in figure 1.

Finally, some estimates of the total number of pages and each page's points were given.

Acknowledgments

The authors would like to thank Narges Abdi for her precious effort in some of simulations.

References

[1] E. Barkan, E. Biham, N. Keller, "Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication", Journal of Cryptology, Volume 21, Number 3, pp 392-429, July 2008.

[2] A. Biryukov, A. Shamir, D. Wagner, "Real Time Cryptanalysis of A5/1 on a PC", Advances in Cryptology, proceedings of Fast Software Encryption'00, LNCS 1978, pp. 1–18, Springer-Verlag, 2001.

[3] J. Golic, "Cryptanalysis of Alleged A5 Stream Cipher", Advances in Cryptology, proceedings of Eurocrypt'97, LNCS 1233, pp. 239–255, Springer-Verlag, 1997.

[4] E. Biham, O. Dunkelman, "Cryptanalysis of the A5/1 GSM Stream Cipher", presented by INDOCRYPT 2000 , LNCS 1977, pp. 43–51, Springer-Verlag, 2000.

[5] P. Ekdahl, T. Johansson, "Another Attack on A5/1", IEEE Transactions on Information Theory, Volume 49, Issue 1, pp. 284-289, 2003.

[6] A. Maximov, T. Johansson, S. Babbage, "An Improved Correlation Attack on A5/1", proceedings of SAC'04, LNCS 3357, pp. 1–18, Springer-Verlag, 2005.

[7] E. Barkan, E. Biham, "Conditional Estimators: An Effective Attack on A5/1", proceedings of SAC' 05, LNCS 3897, pp. 1–19, Springer-Verlag, 2006.

[8] W. G. Chambers, "On random mappings and random permutations," in Fast Software Encryption—Leuven'94 (Lecture Notes in Computer Science), B. Preneel, Ed. Berlin, Germany: Springer-Verlag, 1995, vol. 1008, pp. 22–28.