A SYSTEMATIC ATTACK ON CLOCK CONTROLLED CASCADES

Renato Menicocci

Fondazione Ugo Bordoni
Via B. Castiglione,'59
00142 Roma, Italy

Fax: +39 6 5480 4403
Email: cripto@itcaspur.bitnet

Abstract. Cascades of clock controlled shift registers play an important role in the design of pseudorandom generators for stream cipher cryptography. In this paper, an attack for breaking a kind of such cascades is presented.

1. INTRODUCTION

A technique for obtaining non linear effects in the output sequence of shift register-based binary keystream generators consists in clocking the registers irregularly. A good survey of the structures using irregular clocking has been published by Gollmann and Chambers [1].

One of these structures, commonly referred to as "m-sequence cascade" (see [1] and [2]) or, less properly, as "Gollmann cascade" (see [3], [4], [5]) has aroused a great interest because of its good properties [1]. In spite of this, an intrinsic weakness of such a generator, when *stop-and-go* clocked [1], has been recently pointed out in [5].

In this paper we show how to exploit such a weakness for devising a systematic attack on stop-and-go m-sequence cascades.

2. BACKGROUND MATERIAL AND PRELIMINARY RESULTS

A stop-and-go m-sequence cascade of length L consists of L Linear Feedback Shift Registers (LFSRs), with primitive feedback polynomials of the same degree d, connected as shown in Fig. 1. The first register of the cascade (LFSR 1) is regularly clocked, whereas the clock of the n-th register ($2 \leq n \leq L$) is controlled by the $n-1$ preceding registers. If $s_n^*(t)$ and $e_{n-1}(t)$ are the output of LFSR n and the "clock enable" at the input of AND $(n-1)$, respectively, at the step t, then $e_{n-1}(t) = 0 \Rightarrow s_n^*(t+1) = s_n^*(t)$. Denoting by $\oplus$ the mod 2 addition (XOR operation) we have $e_n(t) = e_{n-1}(t) \oplus s_n^*(t)$, so that the output of the cascade at the step t is given by $e_{L-1}(t) \oplus s_L^*(t)$. In this paper we shall always consider cascades of the described kind.

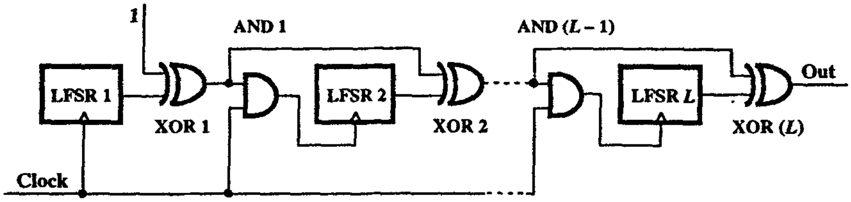

Fig. 1. An L-stage cascade.

The cryptographic interest of this cascade mainly lies in its modularity and in the high values of the *period* (T) and *linear complexity* (LC) of its output sequence. It is known [3] that $T = (2^d - 1)^L$ and $LC \geq d\,(2^d - 1)^{L-1}$.

In spite of these good properties, *short* cascades have been shown to be insecure. In fact, attacks for the cases $L = 2$ and 3 have been proposed (see [4] and [6]). Moreover, an intrinsic weakness of cascades of any length has been studied in [5].

Following [5], to point out the intrinsic weakness of the cascade of Fig. 1, let us suppose that it is driven by L Binary Random Generators (BRGs). We then obtain the model of Fig. 2, where

(i) $G_h^* = \{g_h^*(t)\}$ ($1 \leq h \leq L$) is the sequence generated by BRG h under the control of the sequence $Z_{h-1} = \{z_{h-1}(t)\}$ ($z_0(t) = 1$, $z_k(t) = z_{k-1}(t) \oplus g_k^*(t)$, $1 \leq k \leq L$) ;

(ii) if $h_1, \ldots, h_N$ ($2 \leq h_1 < \ldots < h_N \leq L$) are the values of h such that $z_{h-1}(t) = 0$, then the generators BRG $h_1, \ldots,$ BRG h_N cannot change their output at the next step, that is

$$g_h^*(t+1) = g_h^*(t) \ , \ h \in \{h_1, \ldots, h_N\} \ ;$$

whereas the remaining generators behave, at the step $t+1$, like $L - N$ unconnected BRGs.

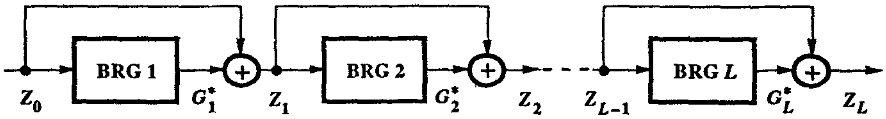

Fig. 2. The cascade model.

In the sequel, with reference to any binary sequence $A = \{a(t)\}$, we will use the following notations

(2.1) $\tilde{A} = \{\tilde{a}(t)\} = \{a(t) \oplus a(t+1)\}$,

(2.2) $A^{(q)} = a(0), a(1), \ldots, a(q-1)$, $q \geq 1$.

In [5] it has been proved that the sequence $\tilde{Z}_L$ can be viewed as the sequence $\tilde{Z}_1$ (see Fig. 2 and (2.1)) corrupted by the *noise* sequence $N_1 = \{n_1(t)\}$ which is generated by a Binary Memoryless Source (BMS) according to the model of Fig. 3. The coincidence probability between $\tilde{z}_1(t)$ and $\tilde{z}_L(t)$ (or, alternatively, the probability of $n_1(t) = 0$) has been shown to be $1/2 + 1/2^L$. In [5] it is also shown how this result can be utilized for determining the sequence at the output of the first register of the cascade of Fig. 1.

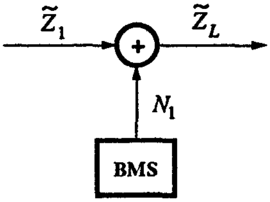

Fig. 3. Correlation between $\tilde{Z}_1$ and $\tilde{Z}_L$.

In the sequel we shall show how the preliminary results established in [5] can be viewed as the first step of a systematic procedure for breaking the entire cascade.

3. FURTHER RESULTS

Theorem 1. The sequence $\tilde{Z}_L$ can be viewed as the sequence $\tilde{Z}_h$ $(h = 1, 2, ..., L-1)$ (see Fig. 2 and (2.1)) corrupted by the *noise* sequence $N_h = \{n_h(t)\}$ which is generated by a BMS. The coincidence probability between $\tilde{z}_h(t)$ and $\tilde{z}_L(t)$ (or, alternatively, the probability of $n_h(t) = 0$) is $1/2 + 1/2^{L+1-h}$.

Sketch of the proof. By the model of Fig. 2 we can derive that the sequence Z_h $(h = 1, 2, \dots, L-1)$ is truly random (see [4] and use induction). Consequently, Z_h can be viewed as the output of the first stage of a cascade of length $L+1-h$ and the model of Fig. 4, where Prob($n_h(t) = 0$) = $1/2 + 1/2^{L+1-h}$, can be obtained [5].

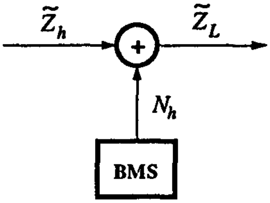

Fig. 4. Correlation between $\tilde{Z}_k$ and $\tilde{Z}_L$.

Denote, for convenience, by $\tilde{Z}_{L,k}$ ($0 \le k \le L-1$) the sequence whose bits, $\tilde{z}_{L,k}(t)$, are given by $\tilde{z}_L(t) \oplus \tilde{z}_k(t)$. By Theorem 1 we then have the following

Corollary 1. The sequence $\widetilde{Z}_{L,h-1}$ can be viewed as the sequence $\widetilde{G}_h^*$ ($1 \le h \le L$) (see Fig. 2 and (2.1)) corrupted by a BMS-generated noise sequence. The coincidence probability between $\tilde{g}_h^*(t)$ and $\tilde{z}_{L,h-1}(t)$ is $1/2 + 1/2^{L+1-h}$.

Sketch of the proof. We simply observe that $\tilde{z}_L(t) = \tilde{z}_h(t) \Leftrightarrow \tilde{z}_{L,h-1}(t) = \tilde{g}_h^*(t)$.

4. CRYPTANALYTIC CONSEQUENCES

Consider now the actual L-stage cascade of Fig. 5, where $S_k^* = \{s_k^*(t)\}$ and $E_k = \{e_k(t)\}$ ($e_0(t) = 1$, $e_k(t) = e_{k-1}(t) \oplus s_k^*(t)$) , $1 \le k \le L$. Denote by $\widetilde{E}_{L,k-1}$ and R_k ($1 \le k \le L$) the sequences whose bits result from $\tilde{e}_{L,k-1}(t) = \tilde{e}_L(t) \oplus \tilde{e}_{k-1}(t)$ and $r_k(t) = \tilde{s}_k^*(t) \oplus \tilde{e}_{L,k-1}(t)$, respectively. Finally, assume that the initial states of the L registers are randomly chosen (the all-zero states are excluded) and that their feedback polynomials $f_1(x), f_2(x), \dots, f_L(x)$ are primitive. The sequences $S_1^*, S_2^*, \dots, S_L^*$ can be then locally modelled by the corresponding sequence $G_1^*, G_2^*, \dots, G_L^*$ of Fig. 2.

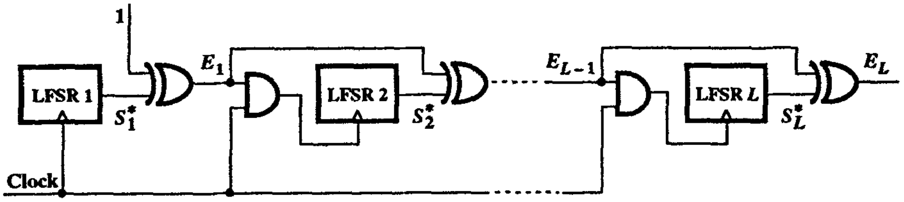

Fig. 5. The actual cascade.

Now, suppose that a sufficiently long segment $E_L^{(N)}$ (see (2.2)) of the output sequence of the cascade is known. In the sequel we shall show that the initial state of the register LFSR h can be then recovered, provided that the segment $E_{h-1}^{(N)}$ and the polynomial $f_h(x)$ are known.

Once $E_L^{(N)}$ and $E_{h-1}^{(N)}$ are given we can obtain the segment $\widetilde{E}_{L,h-1}^{(N-1)}$. On the basis of Corollary 1, we can write

$$(4.1) \qquad \widetilde{E}_{L,h-1}^{(N-1)} = \tilde{s}_h^*(0) \oplus r_h(0), \tilde{s}_h^*(1) \oplus r_h(1), \dots, \tilde{s}_h^*(N-1) \oplus r_h(N-1)$$

$$= s_h^*(0) \oplus s_h^*(1) \oplus r_h(0), s_h^*(1) \oplus s_h^*(2) \oplus r_h(1), \dots, s_h^*(N-1) \oplus s_h^*(N) \oplus r_h(N-1) \ ,$$

where $r_h(t)$ is assumed to be generated by a BMS with

$$(4.2) \qquad \mathrm{Prob}(\ r_h(t) = 0\ ) = 1/2 + 1/2^{L+1-h} \ .$$

Since $E_{h-1}^{(N)}$ is supposed to be known and $e_{n-1}(t) = 0 \Rightarrow s_n^*(t+1) = s_n^*(t)$, it is possible to delete all terms ($s_h^*(t) \oplus s_h^*(t+1) \oplus r_h(t)$) such that $s_h^*(t+1)$ is constrainly equal to $s_h^*(t)$ from the segment $\widetilde{E}_{L,h-1}^{(N-1)}$. By performing this operation we obtain the following segment

$$(4.3)\qquad \Delta^{(M)}_{L,h-1} \overset{\text{def}}{=} \delta_{L,h-1}(0), \dots, \delta_{L,h-1}(M-1)$$

$$= s^*_h(\tau_0) \oplus s^*_h(\tau_0+1) \oplus r_h(\tau_0), \dots, s^*_h(\tau_{M-1}) \oplus s^*_h(\tau_{M-1}+1) \oplus r_h(\tau_{M-1}) \ ,$$

where $\tau_0 < \tau_1 < \dots < \tau_{M-1}$ are the positions of the 1s in the segment $E^{(N-1)}_{h-1}$ and M is, in the average, equal to $N/2$ (observe that, since $e_0(t) = 1$, the segments $\tilde{E}^{(N-1)}_{L,h-1}$ and $\Delta^{(M)}_{L,h-1}$ coincide when $h = 1$). Now, denote by $S^{(M+1)}_h = s_h(0), s_h(1), \dots, s_h(M)$ the segment which would be generated by LFSR h if this register were regularly clocked. Since the register LFSR h is actually clocked only at the steps $\tau_0 < \tau_1 < \dots < \tau_{M-1}$, we easily get

$$(4.4)\qquad \delta_{L,h-1}(t) = s_h(t) \oplus s_h(t+1) \oplus \rho_h(t) = \tilde{s}_h(t) \oplus \rho_h(t), \ 0 \le t \le M-1, \ \rho_h(t) = r_h(\tau_t) \ ,$$

$$(4.5)\qquad \mathrm{Prob}(\ \rho_h(t) = 0\ ) = 1/2 + 1/2^{L+1-h} \ .$$

The equality (4.4) clearly shows that the segments $\Delta^{(M)}_{L,h-1}$ and $\tilde{S}^{(M)}_h$ are correlated. This correlation has been confirmed by several computer simulations. An example of the obtained data is shown in Tab. 1.

$L = 10$, $d = 32$, $N = 1000$, 5000 segments								
h	9	8	7	6	5	4	3	2
mean	0.7499	0.6255	0.5626	0.5313	0.5160	0.5078	0.5040	0.5020
median	0.7500	0.6257	0.5630	0.5312	0.5161	0.5081	0.5040	0.5020
rms	0.7502	0.6259	0.5630	0.5317	0.5165	0.5083	0.5045	0.5025
std. dev.	0.01917	0.02193	0.02223	0.02213	0.02234	0.02237	0.02243	0.02263

Tab. 1. Estimation of $\mathrm{Prob}(\rho_h(t) = 0)$.

Now, since $\tilde{S}^{(M)}_h$ is generable by the register LFSR h [5], we can devise a *correlation attack* [2] for reconstructing the sequence $\tilde{S}_h$ according to the model of Fig. 6. A further step [5] allows us to reconstruct the sequence S_h by using the recurrence relation associated to the feedback polynomial $f_h(x)$ which is supposed to be known.

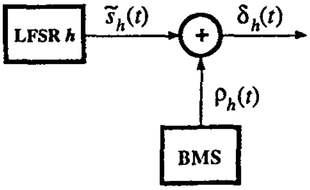

Fig. 6. Model for the correlation attack.

Once the segment $S^{(N)}_h$ has been reconstructed, we can readily get the segments $E^{(N)}_h$ and $\tilde{E}^{(N-1)}_{L,h}$. Thus, the previous procedure can be now used to recover the state of the register LFSR (h + 1) provided that the polynomial $f_{h+1}(x)$ is known, and so on. Since $E^{(N)}_0$ (that is the segment $e_0(0), e_0(1), \ldots, e_0(N-1)$) is always an all-one segment, the first iteration of the procedure ($h = 1$) requires only the knowledge of $E_L^{(N)}$ and $f_1(x)$. It follows that the L-stage cascade of Fig. 5 can be broken by L iterations of the above procedure, provided that the polynomials $f_1(x), f_2(x), \ldots, f_L(x)$ and the segment $E_L^{(N)}$ are known.

5. CONCLUSIONS

The attack presented in this paper is based on the correlation existing between the output sequence of the considered cascade and the output sequences of its intermediate stages. We can easily see that a sufficient condition for the attack to fail is the infeasibilty of its first step. Consequently, since the correlation probability given by (4.5) for $h = 1$ vanishes as L increases, the needed level of practical security can be obtained by making L sufficiently large.

REFERENCES

[1] D. GOLLMANN and W.G. CHAMBERS, "Clock Controlled Shift Registers: A Review" *IEEE J. Selected Areas Commun.*, vol. 7, pp. 525-533, May 1989.

[2] R.A. RUEPPEL. "Stream Ciphers" in *Contemporary Cryptology* (G.J. Simmons, ed.). New York: IEEE Press, 1992.

[3] K.C. ZENG, C.H. YANG, D.Y. WEI and T.R.N. RAO. "Pseudorandom Bit Generators in Stream-Cipher Cryptography" *Computer* 24 , 2 (1991), pp. 8-17.

[4] R. MENICOCCI. "Cryptanalysis of a Two-Stage Gollmann Cascade Generator" *Proc. SPRC '93* (W. Wolfowicz, ed.), pp. 62-69. Rome (Italy), February 1993.

[5] R. MENICOCCI. "Short Gollmann Cascade Generators May Be Insecure". *Proc. Fourth IMA Conf. on Cryptography and Coding*, Cirencester (UK), Dicember 1993 (to appear).

[6] M. MARTURANO. "L'Algoritmo della Sindrome Lineare negli Attacchi Crittanalitici a Correlazione" (in Italian) Laurea Thesis, University "La Sapienza", Infocom Dept., Rome, 1992.