# On the Security of the Gollmann Cascades 

Sang-Joon Park Sang-Jin Lee Seung-Cheol Goh<br>Electronics and Telecommunications Research Institute 161 Kajong-Dong, Yusong-Gu, Taejon, 305-350, Korea<br>e-mail: goh@dingo.etri.re.kr


#### Abstract

The purpose of this paper is to evaluate the security of the Gollmann $m$-sequence cascades of $k$ stages. We give some theoretical results, which can be utilized to construct the transition matrix $T_{n}$ of the conditional probabilities between the input and output strings of a stage. And then, we describe an attack algorithm for guessing the initial state of the first LFSR with desired reliability, using the transition matrix $S_{n}=T_{n}^{k-1}$ of the conditional probabilities between the input string of the second stage and the output of the final stage of the given $k$ stage cascade. We finally evaluate the security of the cascades against this attack. Menicocci recently conjectured that there do not exist the complete analysis of the Gollmann cascades of more than 4 stages and it is infeasible to attack the 10 -stage cascades with LFSRs of degree 100. Our experimental results show that the 9 -stage cascades with LFSRs of degree 100 are completely breakable and the 10 -stage cascades may be insecure.


## 1 Introduction

The purpose of this paper is to evaluate the security of the Gollmann $m$-sequence cascades of $k$ stages[1].

A Gollmann $m$-sequence cascade of $k$ stages consists of a series of $k$ Linear Feedback Shift Registers(LFSRs), with primitive feedback polynomials of same degree $n$. It produces pseudo-random binary sequences of period $\left(2^{n}-1\right)^{k}$, linear equivalence exceeds $n\left(2^{n}-1\right)^{k-1}[1]$. The first LFSR is regularly clocked, whereas all registers except the first are clock controlled by their predecessors. A binary input bit $a_{t}$ clocks a LFSR if $a_{t}=1$, and then is added to the output from the LFSR to give output $c_{t}$, which becomes the input of the next.

$$
\begin{equation*}
c_{t}=a_{t}+b\left(h_{t}\right) \bmod 2, h_{t}=h_{t-1}+a_{t} \bmod 2^{n}-1, t=0,1,2, \cdots \tag{1}
\end{equation*}
$$

with the initial condition $h_{-1}=0$, where $b(\cdot)$ is the sequence generated by the LFSR.

A weakness of the Gollmann cascades, called lock-in effect, was studied in [2]. But the attack by lock-in effect requires $10^{21}$ iterations to analysis the two-stage cascades with polynomials of degree 34. Menicocci proposed an efficient attack on the two-stage Gollmann $m$-sequence cascades, which utilized the correlations between the input and output strings of the final stage of cascades[4]. He also


Fig 1: A stage of the $m$-sequence cascade; The input bit $a_{t}$ clocks LFSR, and then is added to the output from the LFSR to give output $c_{t}$. The delay means the addition takes place after clock.
proposed an attack on the Gollmann cascade of $k$ stages, and conjectured there do not exist algorithms for the complete analysis of the cascade of more than 4 stages, and it is infeasible to attack on the $m$-sequence cascade consists of more than 10 stages with primitive polynomials of degree greater than 100[5].

In this paper, we give some theoretical results, which can be utilized to construct the transition matrix $T_{n}$ of the conditional probabilities between the input and output strings of a stage. And then, we describe an attack algorithm for guessing the initial state of the first LFSR with desired reliability, using the transition matrix $S_{n}=T_{n}^{k-1}$ of conditional probabilities between the input string of the second stage and the output of the final stage of the given $k$-stage cascade. We also evaluate the security of the cascades against this attack. We finally give experimental results on the cryptanalysis of the cascades with less than 10 stages. Our experimental results show that the 9 -stage cascades with LFSRs of degree 100 are completely breakable and the 10 -stage cascades may be insecure.

## 2 Constructions of Transition Matrices

Definition Let $w=x_{0} x_{1} \cdots x_{n-1}$ be a $n$-bit word. Then the value of $w$, denoted by $f(w)$, is

$$
\begin{equation*}
f(w)=\sum_{k=0}^{n-1} x_{k} 2^{n-1-k} \tag{2}
\end{equation*}
$$

Definition Let $w$ and $v$ be n-bit words such that $f(w)=i, f(v)=j$. And let $t^{n}[i, j]$ be the conditional probability that a given stage of the cascade produce $v$ when $w$ is applied to it. Then the $2^{n} \times 2^{n}$ matrix $T_{n}=\left(t^{n}[i, j]\right), 0 \leq i<2^{n}$,
$0 \leq j<2^{n}$, is called a transition matrix of the correlation probabilities between the input and output words of the stage.
Remark For a given cascade of $k$ stages, $S^{n}=\left(T_{n}\right)^{k-1}=\left(s^{n}[i, j]\right)$ is the transition matrix of the conditional probabilities, where $s^{n}[i, j]$ is the probability that the final stage generates $v=f^{-1}(j)$ when $w=f^{-1}(i)$ is applied to the second stage of the given cascade.
Example If $a_{t}=0, c_{t}=b\left(S_{t}\right)$. The sequence $b(\cdot)$ is generated by LFSR, so we have $P\left(c_{t}=0\right)=P\left(c_{t}=1\right)=\frac{1}{2}$. If $a_{t}=a_{t+1}=0$, then $c_{t+1}=b\left(S_{t+1}\right)=$ $b\left(S_{t}\right)=c_{t}$, so we have $P\left(c_{t+1}=c_{t}=0\right)=P\left(c_{t+1}=c_{t}=1\right)=\frac{1}{2}$. Continuing this process, we will finally have

$$
T_{2}=\left(\begin{array}{cccc}
\frac{1}{2} & 0 & 0 & \frac{1}{2} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
0 & \frac{1}{2} & \frac{1}{4} & 0 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}
\end{array}\right),\left(T_{2}\right)^{2}=T_{2} \circ T_{2}=\left(\begin{array}{cccc}
\frac{3}{8} & \frac{1}{8} & \frac{1}{8} & \frac{3}{8} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}
\end{array}\right)
$$

Definition Let $S=\{t, t+1, \cdots, t+n-1\}$ be a set of $n$ consecutive non-negative integers, and let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ be a $n$-bit word such that $f(w)=i$. Then, we may define a class $P_{i}$ of subsets $S_{k}$ by following procedure.

## Procedure 1

Set $k=1$ and $S_{k}=\{t\}$.
For all $l=t+1, \cdots, t+n-1$
If $a_{l}=0$, then set $S_{k}=S_{k} \cup\{1\}$.
Otherwise, set $k=k+1$ and set $S_{k}=\{l\}$.
Property 1 Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ be a $n$-bit word such that $f(w)=i$. And let $m$ be the number of ones in $\left\{a_{t}, a_{t+1}, \cdots, a_{t+n-1}\right\}$. Then the number of subsets $\left|P_{i}\right|$ of the class $P_{i}$ is

$$
\left|P_{i}\right|= \begin{cases}m & \text { if } a_{t}=1  \tag{3}\\ m+1 & \text { otherwise }\end{cases}
$$

Proof. By definition of the class $P_{i}$, it is trivial.
Property 2 Assume that a stage of a cascade produces $v=c_{t} c_{t+1} \cdots c_{t+n-1}$ when $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ is applied. If $P_{i}=\left\{S_{1}, \cdots, S_{m}\right\}$ is the class determined by Procedure 1 , then, for all 1 in $S_{k}$,

$$
\begin{equation*}
a_{l}+c_{l} \bmod 2=\text { const } \tag{4}
\end{equation*}
$$

Proof. From (1), it is trivial.
Property 3 For given integers $i$ and $j$, where $0 \leq i<2^{n}, 0 \leq j<2^{n}$, let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ and $v=c_{t} c_{t+1} \cdots c_{t+n-1}$ be the binary representations of $i$ and $j$ respectively. And let $P_{i}=\left\{S_{1}, \cdots, S_{m}\right\}$ be the class determined by Procedure 1. Then $t^{n}[i, j]=0$ if and only if there exists an integer $0 \leq k \leq m$ such that $a_{l}+c_{l} \neq a_{l^{\prime}}+c_{l^{\prime}} \bmod 2$ for some $l, l^{\prime}$ in $S_{k}$,

Proof. By Property 2, it can be easily proved.
Property 4 If $t^{n}[i, j] \neq 0$ for some $0 \leq i<2^{n}, 0 \leq j<2^{n}$, then $t^{n}[i, j]=\frac{1}{2^{P_{i}} \cdot}$. Proof. Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ be the binary representations of $i$. And let $P_{i}$ be the class determined by Procedure 1. If a stage produces $v=c_{t} c_{t+1} \cdots c_{t+n-1}$ when $w$ is applied, then, by Property 2, there exist constants $b_{1}, b_{2}, \cdots, b_{\left|P_{i}\right|}$, such that $c_{l}=a_{l}+b_{k} \bmod 2$ for all $l$ in $S_{k}$, where $S_{k}$ is in $P_{i}$. This means that, when $w$ is applied, all outputs are uniquely determined by $b_{1}, b_{2}, \cdots, b_{P_{i}}$, so that if $t^{n}[i, j] \neq 0, t^{n}[i, j]=\frac{1}{2^{P_{i}}}$.

Property 5 For a given integer $0 \leq i<2^{n}$, there are exactly $2^{P_{i}}$ number of nonzero elements in $T_{n, i}$, where $T_{n, i}$ is the $i$-th row of the transition matrix $T_{n}$.
Proof. By Property 4, trivial.
Property 6 If $0 \leq j<2^{n-1}, t^{n}[i, j]=t^{n}\left[i, 2^{n}-1-j\right]$ for all $0 \leq i<2^{n}$.
Proof. Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ and $v=c_{t} c_{t+1} \cdots c_{t+n-1}$ be $n$-bit words such that $f(w)=i, f(v)=j$. If the value of $\bar{v}=d_{t} d_{t+1} \cdots d_{t+n-1}$ is $2^{n}-1-j$, then we have, $d_{l}=c_{l}+1 \bmod 2$ for all $l, l=t, \cdots, t+n-1$. Let $P_{i}=\left\{S_{1}, S_{2}, \cdots, S_{\left|P_{i}\right|}\right\}$ be the class determined by Procedure 1. If $t^{n}[i, j]=0$, by Property 3, there is an integer $k, 1 \leq k \leq\left|P_{i}\right|$, such that $a_{l}+c_{l} \neq a_{l^{\prime}}+c_{l^{\prime}} \bmod 2$ for some $l, l^{\prime}$ in $S_{k}$. Hence $a_{l}+d_{l} \neq a_{l^{\prime}}+d_{l^{\prime}} \bmod 2$, so $t^{n}\left[i, 2^{n}-1-j\right]=0$. If $t^{n}[i, j] \neq 0$, by Property 3, we have, $a_{l}+d_{l}=a_{l}+c_{l}+1 \bmod 2=$ const for all $l$ in $S_{k}$. Hence, by Property $4, t^{n}[i, j]=t^{n}\left[i, 2^{n}-1-j\right]=\frac{1}{2^{P_{i i}}}$.

Property 7 For all $0 \leq i, j<2^{n-1}, t^{n}[i, j]=t^{n}\left[2^{n-1}+i, 2^{n-1}+j\right]$.
Proof. We first note that $P_{i}=P_{2^{n-1}+i}$, where $P_{i}$ and $P_{2^{n-1}+i}$ are the classes determined by Procedure 1. Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}, v=c_{t} c_{t+1} \cdots c_{t+n-1}$. Then $w^{\prime}=\tilde{a}_{t} a_{t+1} \cdots a_{t+n-1}$ and $v^{\prime}=\bar{c}_{t} c_{t+1} \cdots c_{t+n-1}$ are the binary representations of $2^{n-1}+i a n d 2^{n-1}+j$ respectively. Since $a_{t}+c_{t}=\overline{a_{t}}+\overline{c_{t}}$, we have $t^{n}[i, j]=t^{n}\left[2^{n-1}+i, 2^{n-1}+j\right]$.
Similarly, we can prove:
Property 8 For all $0 \leq i, j<2^{n-1}, t^{n}\left[2^{n-1}+i, j\right]=t^{n}\left[i, 2^{n-1}+j\right]$.
Property 9 In the case of $n \geq 3, t^{n}[i, j]=0$ if $0 \leq i<2^{n-2}, 2^{n-2} \leq j<2^{n-1}$. Proof. Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ and $v=c_{t} c_{t+1} \cdots c_{t+n-1}$ be $n$-bit words such that $f(w)=i$ and $(v)=j$. Then, we have, $a_{t}=a_{t+1}=c_{t}=0, c_{t+1}=1$. We note that $t, t+1$ belong to same set $S_{1}$ in the class $P_{i}$. But $a_{t}+c_{t} \neq a_{t+1}+c_{t+1} \bmod 2$ so $t^{n}[i, j]=0$.
Property 10 For all $0 \leq i, j<2^{n-2}, t^{n}[i, j]=t^{n-1}[i, j]$.
Proof. Let $P_{i}$ be the class of subsets of the set $S=\{t, t+1, \cdots, t+n-1\}$, and let $P_{i}^{\prime}$ be the class of subsets of the set $S=\{t+1, \cdots, t+n-1\}$ determined by Procedure 1, when $i$ is applied. Note $\left|P_{i}\right|=\left|P_{i}^{\prime}\right|, S_{1}^{\prime}=S_{1}-\{t\}, S_{k}^{\prime}=S_{k}, k>1$. Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ and $v=c_{t} c_{t+1} \cdots c_{t+n-1}$ be the binary representations of $i, j$, respectively. If $t^{n}[i, j] \neq 0$, then $a_{l}+c_{l} \bmod 2=$ const if $l$ in $S_{k}$ for all $S_{k}$ in $P_{i}$. Hence $a_{l}+c_{l} \bmod 2=$ const if $l$ in $S_{k}^{\prime}$ for all $S_{k}^{\prime}$ in $P_{i}^{\prime}$, so that $t^{n-1}[i, j] \neq 0$.

Therefore, if $t^{n}[i, j] \neq 0$, then, by Property $4, t^{n}[i, j]=t^{n-1}[i, j]=\frac{1}{2^{P_{i}} .}$. In the case of $t^{n}[i, j]=0$, we can prove similarly $t^{n}[i, j]=t^{n-1}[i, j]=0$.

Property 11 For all $2^{n-2} \leq i<2^{n-1}, t^{n}[i, j]=\frac{1}{2} t^{n-1}[i, j]$.
Proof. Let $P_{i}$ and $P_{i}^{\prime}$ be the classes defined in the proof of Property 10. In a similar way, we can easily prove that $t^{n}[i, j]=0$ if and only if $t^{n-1}[i, j]=0$. We note that, by Property $1,\left|P_{i}\right|=\left|P_{i}^{\prime}\right|+1$. Hence, by Property 4, if $t^{n}[i, j] \neq 0$, $t^{n}[i, j]=\frac{1}{2} t^{n-1}[i, j]$.

Property 12 Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$ be the binary representation of $i$. Then $t^{n}[i, 0]=0$ if and only if there exist $l, l^{\prime}$ such that $t \leq l<l^{\prime} \leq t+n-1, a_{l}=1$ but $a_{l^{\prime}}=0$.
Proof. Assume that $a_{l}=1$ but $a_{l^{\prime}}=0$ for some $l, l^{\prime}$ such that $t \leq l<l^{\prime} \leq t+n-1$. Let $f(w)=i$. And let $S_{k}$ be the subset of $P_{i}$ which contains $l^{\prime}$. Let us define $h=\max \left\{m \mid l \leq m<l^{\prime}, a_{m}=1\right\}$. Then $h \in S_{k}$ and $a_{h}+0 \neq a_{l^{\prime}}+0 \bmod 2$ so that $t^{n}[i, 0]=0$. Suppose that $t^{n}[i, 0]=0$. Then it is clear that $i \neq 0,1$. Hence there exists at least one $l$ such that $t \leq l<t+n-1$ and $a_{l}=1$, say $p=\min \left\{l \mid t \leq l<t+n-1, a_{l}=1\right\}$. In the case of $p=t$, if $a_{l}=1$ for all $t+1 \leq l \leq t+n-1$, then each $S_{k}, k=1,2, \cdots, t+n-1$, has only one element so that $t^{n}[i, 0] \neq 0$. So we may assume $p \geq t+1$ ie $a_{t}=0$. If $a_{l}=1$ for all $l$, $p \leq l \leq t+n-1$, then only $S_{0}$ may have more than one elements. Since all $a_{l}$ in $S_{0}$ are 0 so $t^{n}[i, 0] \neq 0$.

From Property 12, we have
Property 13

$$
t^{n}[i, 0]=t^{n}\left[i, 2^{n}-1\right]=\left\{\begin{array}{cl}
\frac{1}{2^{k+1}} & \text { if } i=2^{k}-1 \text { for some } k, 0 \leq k \leq n-1 \\
\frac{1}{2^{n}} & \text { if } i=2^{n}-1 \\
0 & \text { Otherwise }
\end{array}\right.
$$

Now we summarize Properties solved in this section to give Theorem 1.
Theorem 1. For all $n \geq 3$, let $H_{n}=\left(h^{n}[i, j]\right)$, $H_{n}^{\prime}=\left(h^{\prime n}[i, j]\right)$ be $2^{n-1} \times 2^{n-1}$ matrices of the probabilities such that $h^{\prime n}[i, j]=h^{n}\left[i, 2^{n-1}-1-j\right]$. Then 1) the transition matrix $T_{n}=\left(t^{n}[i, j]\right)$ can be represented as:

$$
T_{n}=\left(\begin{array}{ll}
H_{n} & H_{n}^{\prime}  \tag{5}\\
H_{n}^{\prime} & H_{n}
\end{array}\right)
$$

2) the $2^{n-1} \times 2^{n-1}$ matrix $H_{n}=\left(h^{n}[i, j]\right)$ can be represented as:

$$
H_{n}=\left(\begin{array}{cc}
H_{n-1} & 0  \tag{6}\\
\frac{1}{2} H_{n-1}^{\prime} & \frac{1}{2} H_{n-1}
\end{array}\right)
$$

with the initial condition $H_{1}=\left(\frac{1}{2}\right)$.

## 3 The attack proposed

Our aim is to find all initial states of LFSRs of the given cascade, under the assumption that we know a sufficiently large number of consecutive bits of the output and all primitive polynomials.
For a given cascade of $k$ stages, we define a matrix $S^{n}$ by

$$
\begin{equation*}
S^{n}=\left(T_{n}\right)^{k-1}=\left(s^{n}[i, j]\right) \tag{7}
\end{equation*}
$$

Then the component $s^{n}[i, j]$ is the probability that the final stage produces $v=f^{-1}(j)$ when $w=f^{-1}(i)$ is applied to the second stage of the cascade. This matrix $S^{n}$ enables us to guess the initial state of the LFSR of the first stage with desired reliability.

Theorem 2. For all $i, 0 \leq i<2^{n}$, let $w_{i}$ be the $n$-bit word such that $f\left(w_{i}\right)=i$, and $A_{i}$ the event that $w_{i}$ is applied to the second stage of the cascade. And let $E_{0}$ be the event that the cascade generates a run of $n$ consecutive zeros and $E_{1}$ a run of $n$ consecutive ones. If $E=E_{0} \cup E_{1}$, then

$$
\begin{equation*}
P\left(A_{i} \mid E\right)=s^{n}[i, 0] \tag{8}
\end{equation*}
$$

where $P\left(A_{i} \mid E\right)$ denotes the conditional probability of $A_{i}$ occurs given that $E$ has occurred.

Proof. By Bayes's theorem, we have,

$$
\begin{aligned}
P\left(A_{i} \mid E_{0}\right) & =\left\{P\left(A_{i}\right) P\left(E_{0} \mid A_{i}\right)\right\} /\left\{\sum_{m=0}^{2^{n}-1} P\left(A_{m}\right) P\left(E_{0} \mid A_{m}\right)\right\} \\
& =P\left(E_{0} \mid A_{i}\right) / \sum_{m=0}^{2^{n}-1} P\left(E_{0} \mid A_{m}\right)=s^{n}[i, 0] / \sum_{m=0}^{2^{n}-1} s^{n}[m, 0]=s^{n}[i, 0]
\end{aligned}
$$

Hence, $P\left(A_{i} \mid E\right)=\frac{1}{2} P\left(A_{i} \mid E_{0}\right)+\frac{1}{2} P\left(A_{i} \mid E_{1}\right)=\frac{1}{2}\left(s^{n}[i, 0]+s^{n}\left[i, 2^{n}-1\right]\right)=s^{n}[i, 0]$.
Corollary 3. Let $w=a_{t} a_{t+1} \cdots a_{t+n-1}$. Then, for all $k, 0 \leq k \leq n-1$,

$$
\begin{equation*}
P\left(a_{t+k}=0 \mid E\right)=\sum_{i} s^{n}[i, 0] \tag{9}
\end{equation*}
$$

where $i$ ranges over integers such that $i \wedge 2^{n-k-1}=0$. The notation $\wedge$ denotes the bitwise $A N D$ of two integers. Hereafter, for all $0 \leq k \leq n-1$, we define $q_{k}$ by $q_{k}=1-P\left(a_{t+k}=0 \mid E\right)$.

For a precise description of our attack, let $m$ be the same degree of primitive polynomials and let $p$ be the desired error rate. In order to find the initial state of LFSR in the first stage, we first scan the given output bits to find all runs of at least $n$ consecutive zeros or ones. Then we scan each run for $k$ such that $q_{k} \leq p$, to set linear equations, which have the $m$ unknowns for the initial state.

In this way, we can set a sufficiently large number, say $l$, of linear equations. Choose randomly $m$ out of the $l$ equations and solve the linear system. If it has a solution, say $\bar{x}=x_{0}, x_{1}, \cdots, x_{m-1}$, then examine how many equations hold when $\bar{x}$ is substituted. If it is the real initial state, the number of equations held is on average $l(1-p)$. This is repeated for each except final stage. And then, apply the algebraic technique[6] to the final stage to guess the initial state of its LFSR. Test, finally, whether the cascade produces the given output sequence. If not, repeat again.

How many trials of selections $m$ from $l$ equations is needed to find the initial state of the first stage. By a similar way in [3], we can estimate the probability $q$ that a trial successes, so $q^{-1}$ is the expected number of trials.

$$
\begin{equation*}
q=\left(1-\frac{p l}{l}\right)\left(1-\frac{p l}{l-1}\right) \cdots\left(1-\frac{p l}{l-m+1}\right) \geq\left(1-\frac{p l}{l-m+1}\right)^{m} \tag{10}
\end{equation*}
$$

Example Consider a cascade of 2 stages with LFSRs of the same length 34. We are assumed to know 800 consecutive output bits, so that we can find, on average, 25 number of runs whose lengths are greater than or equal to 5 . If we put $p=2^{-4}$, we can set 50 linear equations so $q \approx 10^{-3}$. Hence about 1000 trials are expected to find the correct solution.

Algorithm: An attack on the Gollmann k-stage cascades
Step 1 Scan the given output bits, $a_{h}, a_{h+1}, a_{h+2}, \cdots \cdots$, , to find all runs of at least $n$ consecutive zeros or ones. This number $n$ may be determined by the desired error rate $p$, the number of given output bits, and the degrees of LFSRs.
Step 2 Repeat the process of Steps 3-6 in order to guess the initial states of LFSRs of all but final stages with desired reliability $1-p$.
Step 3 Select $a_{t}$ such that $P\left(a_{t}=0 \mid E\right) \geq 1-p$ to set a linear equation, which has the $m$ unknowns for the initial state. In this way, we can set a sufficiently large number, say $l$, of linear equations.
Step 4 Select randomly $m$ out of the $l$ equations and solve the linear system.
Step 5 If it has no solution, goto Step 4.
Step 6 Examine how many equations hold when the solution is substituted. If the number is less than $k$, which is heuristically determined by $l(1-p)$, goto Step 4.
Step 7 Apply the algebraic technique to the $k$ th stage using the sequence generated by the $k-1$ th stage.
Step 8 Test whether the cascade produces the correct final output sequences. If success terminate, else goto Step 2.

## 4 Security evaluation

The security of the Gollmann cascades may be evaluated by the number of trials for obtaining the correct initial state of its first LFSR and the number of the final output bits for obtaining runs whose lengths are sufficiently large enough
to guarantee small error rate. From the view point of cryptanalysis, there may be trade-off between the number of trials and bits. When the length of runs is greater than 9 , we could not calculate correctly the error rate and the number of trials due to the memory problem of computer. However, by some statistical observation, we expect that the runs of length 20 enable us to break completely the cascades of 10 stages with LFSRs of degree 100.

We have implemented the proposed attack in C on an Axil Hyundai workstation( 80 MIPS), compatible with a SUN Sparc 10, in a UNIX environment. For the case of 2 -stage cascades with primitive polynomials of degree 34 , we have found all initial states of LFSRs within 2.8 CPU seconds on average.
Following table gives our experimental results. In those tables, Degree denotes the degree of primitive polynomials, Stage the numbers of stages, Bit the numbers of output bits, Run the lengths of runs, $p$ desired error rates, Eqns the numbers of equations, Trial the numbers of trials, Theo the numbers theoretically calculated, Expe the numbers of experimental results, CPU the CPU seconds needed execution the algorithm.

Table: Experimental Results

| Degree | Stage | Bit | Run | $p$ | Eqns | Trial |  | CPU |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | Theo | Expe |  |
| 34 | 2 | 800 | 5 | 0.0625 | 75 | 18 | 15 | 2.8 |
|  | 3 | 1800 | 6 | 0.1269 | 82 | 423 | 119 | 4.5 |
|  | 4 | 7000 | 8 | 0.1208 | 73 | 399 | 96 | 7.2 |
|  | 5 | 35000 | 10 | 0.1145 | 110 | 146 | 241 | 49.4 |
| 44 | 2 | 1000 | 5 | 0.0625 | 95 | 46 | 23 | 8.7 |
|  | 3 | 2300 | 6 | 0.1269 | 116 | 2033 | 626 | 17.4 |
|  | 4 | 9000 | 8 | 0.1208 | 109 | 1576 | 436 | 28.4 |
|  | 5 | 45000 | 10 | 0.1145 | 117 | 891 | 255 | 68.8 |
| 54 | 2 | 1200 | 5 | 0.0625 | 115 | 114 | 437 | 18.2 |
|  | 3 | 3000 | 6 | 0.1269 | 127 | 16734 | 1843 | 56.7 |
|  | 4 | 15000 | 8 | 0.0801 | 136 | 327 | 873 | 99.2 |
|  | 5 | 60000 | 10 | 0.0801 | 121 | 412 | 271 | 155.3 |
| 100 | 3 | 20000 |  | 0.04699 | 304 | 350 | 896 | 123.0 |
|  | 4 | 60000 | 9 | 0.05235 | 256 | 949 | 1892 | 265.1 |
|  | 5 | 240000 | 11 | 0.05318 | 263 | 1006 | 3950 | 687.3 |
|  | 6 | 500000 | 12 | 0.08440 | 242 | 99874 | 4282 | 721.0 |
|  | 7 | 2000000 | 14 | 0.08960 | 230 | 267088 | 1741 | 828.0 |
|  | 8 | 8000000 | 16 | 0.08562 | 262 | 88956 | 39147 | 2334.0 |
|  | 9 | 30000000 | 18 | 0.07188 | 248 | 15340 | 51006 | 3322.0 |

## 5 Concluding Remarks

In this paper, we have evaluated the security of the Gollmann $m$-sequence cascades. We have described, in detail, an algorithm for guessing the initial state of
the first LFSR with desired reliability. We have also given experimental results on the complete analysis of 9 -stage cascades with polynomials of degree 100 . We have finally defined the security of cascades by the number of trials for obtaining the correct initial state of its first LFSR and the number of the final output bits obtaining runs whose lengths are sufficiently large enough to guarantee small error rate. Now we are trying the complete analysis of 10 -stage cascades with LFSRs of degree 100 . It is expected to be successful with $2^{26}$ output bits of the cascades.

## References

1. Gollmann, D.: Pseudorandom Properties of Cascaded Connections of Clock Controlled Shift Registers, Advances in Cryptology-Eurocrypt'84, Lecture notes in Computer Science 209, (1985) 93-98
2. Chambers, W. G., Gollmann, D.: Lock in Effect in Cascades of Clock Controlled Shift Register, Advances in Cryptology-Eurocrypt'88, Lecture notes in Computer Science 330, (1989) 331-343
3. Meier, W., Staffelbach, O.: Correlation Properties of Combiners with Memory in Stream Ciphers, Journal of Cryptology, 5, No. 1, (1992) 67-86
4. Menicocci, R.: Cryptanalysis of a Two-Stage Gollmann Cascade Generator, Proc. the 3rd Symposium of State and Progress of Research in Cryptography, Rome, (1993) 62-69
5. Menicocci, R.: Short Gollmann Cascade Generator May Be Insecure, presented at Fourth IMA Conf. on Cryptography and Coding, (1993)
6. Beker, H., Piper, F,: Cipher Systems: The Protection of Communications, Wileyinterscience, (1982)
