Investigation on Pseudorandom Properties of FCSR Sequence

Yu Zheng, Xiaohu Tang, Dake He, Lixing Xu
School of Computer & Comm. eng., Southwest Jiaotong University, Chengdu, Sichuan 610031, P. R. China
zyu_swjtu@163.com or cdzhengyu@yahoo.com.cn

Abstract—Feedback with Carry Shift Register (FCSR) proposed a new path for pseudorandom sequence design, however its properties still need to be revealed deeply. In this paper compared with LFSR, the properties of FCSR are reviewed briefly and the present conclusions about pseudorandom properties of LFSR are analyzed as well. Both full-period and half-period l-sequences generated by FCSR are tested by eight tests in NIST (National Institute of Standardization and Technology) STS package to investigate its pseudorandom property including balance, uniformity, linear span, linear dependence, runs, compressible property and deBruijn property etc. As a result, both full-period and half-period l-sequences that present the real feature of FCSR sequence in deed, have passed all tests but Discrete Fourier Transform (DFT) test when the length of tested subsequences becomes long enough. That means FCSR with excellent pseudorandom property is a good component of stream cipher generator but can’t be used individually. It is recommended to combine LFSR with FCSR to design stream cipher in the future, which have passed all eight tests in this paper.

I. INTRODUCTION

The pseudorandom sequence design has been a hot topic in the areas of Cryptology and Communication. Generally, Linear Feedback Shift Register (LFSR), combined with nonlinear-filter, nonlinear-combining and clock-controlled model etc. is popular in fast implementation in hardware to generate key stream [1]. An enormous amount of efforts have been directed towards LFSR and m-sequence, whose algebra architecture has also been revealed deeply. So, the attacks to LFSR are more and more mature and perfect.

In 1994, Klapper and Goresky proposed FCSR whose maximal period sequence is called l-sequence [2~4] and analyzed FCSR’s properties in paper [5~10]. Since FCSR is analogous to LFSR in architecture, many parallels between properties of LFSR sequence and those of FCSR sequence are achieved during these years, such as rational approximate algorithm [11], linear span of FCSR [12] and the properties of l-sequence [13].

A good pseudorandom sequence should have good statistical distributions, large period and large linear span etc. The FCSR proposed a new path for generating pseudorandom sequence, but the pseudorandom property of its output sequence, especially for the l-sequence, are still required to be revealed deeply. In paper [14], the pseudorandom properties of the sequences generated by 8-bits FCSR were tested by NIST STS package. But the period of tested sequences are not more than 509 bits that can only satisfy the requirement of frequency test, runs test and the cumulative sums test in STS. Furthermore, they didn’t take the fact into account that the second half of a period of l-sequence is the bitwise complementary of the first half [13]. So the limited testing results on full-period l-sequence in [14] can’t present the pseudorandom properties of l-sequence in detail.

The rest paper is organized as follows: firstly, compared with LFSR, the properties of FCSR are reviewed briefly and the present conclusions about pseudorandom properties of FCSR are analyzed as well. Secondly, eight tests are employed to investigate the pseudorandom property of full-period and half-period l-sequence. So it is beneficial to deep the theoretical basis and present the properties of FCSR, which contributes to development of FCSR and pseudorandom sequence design. Finally the paper is concluded in section 4 and some useful upper and lower bound of strong 2-prime [12] are listed in appendix for FCSR.

II. FCSR ADN ITS MAIN PROPERTIES

As shown in Fig.1, the FCSR with connection integer q is a feedback register with r bits of storage plus additional memory containing an integer for carry. The coefficients $a_i \in GF(2), i = 1, 2, ..., r$ may be thought as taps as shift register. If the content of the register at any given time are $a_n, a_{n-1}, ..., a_{n-r+1}, a_{n-r} \in GF(2)$ and the memory is $m_{n-1} \in N$, the operation of the shift register is defined as follows:

A1 Form an integer sum $\sigma = \sum q_i a_{n-i} + m_{n-1}$.

A2 Shift the content on step to the right, while outputting the rightmost bit $a_{n-r}$.

A3 Put $a_{n} = \sigma \mod 2$ into the leftmost cell of the shift register.

A4 Replace the memory integer $m_{n-1}$ with $m_{n} = (\sigma - a_{n})/2$.

Figure 1. Feedback with carry shift register
Theorem 1 [2]: There is a one-to-one correspondence between rational number $\alpha = p/q$ and eventually periodic binary sequence $a$, which associates such rational number $\alpha$ with the bit sequence $(a_0, a_1, a_2, \ldots)$ of its 2-adic expansion. I.e. $\alpha = \sum_{i=0}^{\infty} a_i 2^i$. Where connection integer $q = \sum_{i=0}^{\infty} q_i 2^i - 1$ is an odd rational number. The sequence $a$ is strictly periodic if and only if $|\alpha| < 1$.

Definition 1 [2]: $q$ is a prime number and 2 is primitive root modulo $q$, then $q$ is 2-prime. If $q=2p+1$, both $p$ and $q$ are 2-prime, then $q$ is called strong 2-prime.

Theorem 2 [2]: An $l$-sequence is a FCSR sequence with maximal possible period $T = \phi(q)$. The best we can hope is that the period is $q-1$ and this occurs when $q$ is 2-prime.

Theorem 3 [2]: Let $\alpha = (a_0, a_1, a_2, \ldots)$ be a periodic sequence generated by FCSR with connection integer $q$. $q$ is a power of a prime number $p$, $q = p^r$, such that 2 is a primitive root modulo $q$, then we have $a_{i+n} + a_i = 1$ for all $i = 0, 1, 2, \ldots$, where $T = \phi(q) = \phi(p^r) = p^r/(p-1)$.

Definition 2 [5]: de Bruijn property is that the numbers of occurrences of any two subsequences in sample sequence differ by at most one.

Theorem 4 [4]: Let $q = p^r$ be the connection integer of an FCSR and $p$ is a prime number. Such that 2 is a primitive root modulo $q$, then the numbers of occurrences of any two subsequences in a single period differ by at most two.

Theorem 5 [12]: If the connection integer $q$ of an FCSR is 2-prime, then the linear span of the FCSR is less than or equals to $(q+1)/2$. If $q = 2p+1$ is a strong 2-prime, then the linear span is $p+1$.

In terms of theorem 1 and 2, we can know that the period of $l$-sequence generated by FCSR is not less than that of $m$-sequence generated by LFSR with same stage. According to theorem 3, the number of zeros is equal to that of ones in one period of $l$-sequence since the second half of a period of $l$-sequence is the bitwise complementary of the first half. Therefore, the half-period sequence or smaller subsequence in $l$-sequence should be tested in order to investigate the balance and uniform properties of $l$-sequence deeply. Meanwhile, theorem 4 indicates that $l$-sequence nearly has the de Bruijn property, which is expected to present perfect results when tested by de Bruijn property test. However, the properties of subsequence are still required to be checked. As shown in theorem 5, linear span of the sequence generated by FCSR whose $q$ is stronger 2-prime is more than that of $m$-sequence generated by LFSR with same stages. The lower and upper bound of strong 2-prime for FCSR stages are less than 28 are searched by computer and listed in appendix, which benefits to choose $q$ for FCSR to generate sequence with large linear span.

As shown in table 1, some important properties of FCSR are compared with that of LFSR, which leads to parallel research on FCSR to LFSR.

<table>
<thead>
<tr>
<th>Properties</th>
<th>LFSRs</th>
<th>FCSRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation Architecture</td>
<td>Galois and Fibonacci</td>
<td>Galois and Fibonacci</td>
</tr>
<tr>
<td>Maximal period sequence</td>
<td>$m$-sequence</td>
<td>$l$-sequence</td>
</tr>
<tr>
<td>Algorithm to compute initial load</td>
<td>Berlekamp-Massey Algorithm</td>
<td>Rational Approximate Algorithm</td>
</tr>
<tr>
<td>Correlation</td>
<td>Periodic correlation</td>
<td>Arithmetic correlation</td>
</tr>
<tr>
<td>Complexity</td>
<td>Linear span</td>
<td>2-adic span</td>
</tr>
<tr>
<td>Algebra expression</td>
<td>Polynomial</td>
<td>$p$-adic number</td>
</tr>
<tr>
<td>Mathematics tool</td>
<td>Finite filed</td>
<td>$p$-adic theory</td>
</tr>
</tbody>
</table>

III. Testing on Pseudorandom Property of FCSR Sequence

Theorem 6 [4]: The rational number $\alpha = p/q$ is a quotient of two integer and the denominator $q$ is the connection integer of an FCSR that generates the sequence $a$. The sequence $a$ is strictly periodic if and only if $|\alpha| < 1$. The value of the memory must lie in the range $0 \leq m \leq wt(q+1)$. Where $wt(x)$ is the hamming weight of $x$. When $m = 0$, the output sequence isn’t degenerate except the initial loadings are all 0.

In order to test the pseudorandom property of the sequence generated by FCSR with different stage and compare the testing results with those in [14], using computer, we search three groups of connection integer $q$ (see table 2) in every of which ten continuous 2-prime numbers are comprised respectively. At the same time, we let all initial value of additional memory in FCSR equal to 0 to ensure the outputted FCSR sequence is strictly periodic according to theorem 6.

As shown in table 3, eight typical tests in STS have been selected to test three groups of FCSR sequences with the connection number $q$ in table 2. The required length $n$ of tested sequence for every test is listed in the last column of table 3. We check the sequences generated by the 1rd group of $q$ separately using test 1-test 4 so as to compare with the testing result in [14]. Meanwhile the sequences generated by the 2nd group of $q$ are tested respectively in use of test 5 and 6. Finally, the sequences generated by the 3rd group of $q$ are tested using all eight tests.

In NIST, testing result is $P$-value $\in [0,1]$. If $P$-value $\geq 0.01$ the tested sequence is considered to be
random and the bigger the \textit{P-value} is, the better pseudorandom property the tested sequence has.

Fig.2-fig.5 show that the full-period and half-period sequences generated by the 1\textsuperscript{st} group of \(q\) have passed test 1-test 4 since all \(P\)-value > 0.01. However, all testing results on half-period \(l\)-sequences are less than those on full-period \(l\)-sequence, which present the real feature of \(l\)-sequence. Furthermore, in Fig.2 and fig.5, all testing results on full-period \(l\)-sequence generated by every \(q\) are 1, which confirm theorem 3 and theorem 4 in turn. That means the full-period sequence has perfect balance of zeros and ones and deBruijn properties.

As shown in fig.6, all full-period and half-period \(l\)-sequences generated by the 2\textsuperscript{nd} group of \(q\) have passed test 5. I.e. the subsequences in \(l\)-sequence have little linear dependence. While Fig.7 indicates that the testing result of DFT is decreased down to 0 little by little when the length of tested sequence becomes long enough. Though there are some anomalous points in the testing results, the tendency to decrease with the increasing length of tested sequence is obvious. In order to explain the meaning of the testing result deeply, the principle of the DFT test is depicted briefly:

(1) The one and zero in tested sequence \(X\) are converted to \(2x_i = -1\). E.g., \((1,0,0,1,1) \rightarrow (-1,-1,-1,1,1)\).

(2) Apply the DFT on \(X\) to produce: 

\[
SDFT_X = S.
\]

Then calculate 

\[
|S'| = |S| \quad \text{where } S \text{ is the first half part of } S.
\]

(3) Compute \(T = \sqrt{3n}\) as the theoretical threshold peak.

\begin{table}[H]
\centering
\caption{Connection integer of FCSR}
\begin{tabular}{cccc}
\hline
\(q\) & 1(stage=8) & 2(stage=16) & 3(stage=20) \\
\hline
\(q_1\) & 373 & 100333 & 2001477 \\
\(q_2\) & 379 & 100357 & 2000221 \\
\(q_3\) & 389 & 100363 & 2002227 \\
\(q_4\) & 419 & 100379 & 2000269 \\
\(q_5\) & 421 & 100403 & 2000309 \\
\(q_6\) & 443 & 100411 & 2000371 \\
\(q_7\) & 461 & 100439 & 2000387 \\
\(q_8\) & 467 & 100469 & 2000389 \\
\(q_9\) & 491 & 100493 & 2000413 \\
\(q_{10}\) & 509 & 100501 & 2000429 \\
\hline
\end{tabular}
\end{table}
Under an assumption of randomness, 95% of the values obtained from the test should not exceed $T$.

(4) Compute the theoretical number of peaks $N_0 = 0.95n/2$ that are less than $T$ and compute the number of actual observed number of peaks $N_1$ in $M$ that are less than $T$.

\[
d = \frac{N_1 - N_0}{\sqrt{0.95 \times 0.05n/2}}, \quad \text{P-value} = \text{erf} \left( \frac{d}{\sqrt{2}} \right),
\]

Where \( \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{-z}^{z} e^{-\omega^2} \, d\omega \).

In terms of equation (5) above, the smaller the $P$-value is, the difference between $N_0$ and $N_1$ is, which contravenes the assumption of randomness of sequence. The reason for the decrease of $P$-value is in that $l$-sequence begins to show repetition to some extent when the length of sequence becomes long.

Some other $l$-sequences with more larger period have been checked by DFT test and the testing results accord with fig.7 and fig.8. So, FCSR with excellent pseudorandom property is a good component of key stream generator but can’t be used individually. It is recommended to use the combination of LFSR and FCSR [16] (see Fig.9), which is expected to have large confusion and pseudorandom property, since addition with carry destroys the algebraic properties of LFSR and the XOR destroys the algebraic properties of FCSR. The scheme proposed in [16] was also examined by all 8 tests with the parameters defined as follows. Where the connection polynomials of LFSRs are all primitive polynomials and the initial loading of additional memory are all zero.

\[f_1(x) = x^{13} + x^7 + x^3 + x + 1, \quad f_2(x) = x^{17} + x^{13} + x^7 + x^3 + 1, \quad f_3(x) = x^{19} + x^{13} + x^5 + x^3 + 1, \quad f_4(x) = x^{19} + x^{17} + x^7 + x^3 + 1\]

$q_1 = 4993; q_2 = 1549; q_3 = 857; q_4 = 509$.

As shown in fig.10, even if the length of tested sequence is increased to $1.2 \times 10^6$ bits, $P$-value is still much larger than 0.01, and we can’t see any evidence of decline, which infers that the sequence has good periodic property. Meanwhile, the sequence generated by this scheme with the parameters above has passed all 8 tests (other testing results aren’t present here in view of the limited length of paper). Therefore the sequences generated by this generator have good pseudorandom properties.

As shown in Fig.11, the results of Lempel-Ziv test on the 3rd group of $q$ are all 1, which illustrates that both full-period and half-period $l$-sequence are uncompressible using Lempel-Ziv Algorithm, which accords with the assumption of randomness. Meanwhile, all $P$-value > 0.01 in fig.11, which infers the linear span of $l$-sequence is large enough to accord with assumption of randomness. Where the results of
linear complexity test on half-period $l$-sequence is less than those of full-period $l$-sequence excluding $q=2000413$. The results of other 6 tests is similar with the testing results on the 1st and 2nd groups of $q$, and don’t shown again for the limited length of the paper. So, the 3rd group of $q$ have passed all tests except for DFT test when the length of tested sequence exceed one tenth of period.

![Figure 10 DFT test on the combination of LFSR and FCSR](image)

![Figure 11 Results of linear complexity test and Lempel-Ziv test](image)

IV. CONCLUSION

In this paper both full-period and half-period $l$-sequences are investigated by eight tests in NIST STS package to show the pseudorandom property of FCSR sequence including balance, uniformity, linear span, linear dependence, runs, compressible property and deBruijn property. As a result, both full-period and half-period $l$-sequences have passed all the tests but DFT test and the testing results on half-period $l$-sequences present the real feature of FCSR sequence in deed. With the increase of the length of $l$-sequence, repetition is shown materially before full period is formed. According to the testing results of DFT, the available length of $l$-sequence for key stream is less than one tenth of a full period. Increase in connection integer $q$ of FCSR has less benefit in the improvement of periodic property and adds hardware resource and computing time in turn. Thus, FCSR with excellent pseudorandom property is a good component of stream cipher but can’t be used individually. It is recommended to use the combination of LFSR with FCSR who have passed all 8 tests in this paper and proposed a new path for pseudorandom sequence design in future.

ACKNOWLEDGMENT

This work is supported by foundation for author of excellent Ph.d thesis under Grant No.200341 and Sichuan youth science foundation under Grant No.04ZQ026-048.

REFERENCES


Appendix

<table>
<thead>
<tr>
<th>Stage</th>
<th>Strong 2-prime</th>
<th>Stage</th>
<th>Strong 2-prime</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
<td>Lower Bound</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>/</td>
<td>/</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>59</td>
<td>59</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>107</td>
<td>107</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>/</td>
<td>/</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>347</td>
<td>347</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>587</td>
<td>1019</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>1307</td>
<td>1307</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>2459</td>
<td>3947</td>
<td>24</td>
</tr>
<tr>
<td>12</td>
<td>4139</td>
<td>6827</td>
<td>25</td>
</tr>
<tr>
<td>13</td>
<td>8699</td>
<td>16187</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>17387</td>
<td>32603</td>
<td>27</td>
</tr>
</tbody>
</table>

TABLE IV. LOW AND UPPE R BOUND OF STRONG 2-PRIME FOR FCSR