Ziv-Lempel Complexity for Periodic Sequences and its Cryptographic Application

Sibylle Mund
Siemens AG
ZFE IS KOM4
Otto-Hahn-Ring 6
8000 München 83
West Germany

Abstract

The Ziv-Lempel complexity is a well-known complexity measure. In our paper we consider the Ziv-Lempel complexity for periodic sequences as well as for pseudorandom number sequences. Further on, we will look at its cryptographic significance and compare it with other complexity measures such as the linear complexity.

1 Introduction

In the last couple of years several different complexity measures were used to examine pseudorandom number sequences in cryptography. Examples for such complexity measures are the linear complexity which is defined in Rueppel [Ruep 86] or the maximal-order complexity which was introduced by Jansen [Jans 89]. Both complexity measures can be used to test pseudorandom number sequences against the qualities of random number sequences has and therefore to distinguish between pseudorandom number sequences with good qualities and those with bad qualities.

Another co nplexity measure for sequences was defined by Ziv and Lempel in 1976 [Lemp 76]. This complexity measure is a measure of the rate at which new patterns emerge as we move along the sequence. Until now the Ziv-Lempel complexity was mainly used in connection with the Ziv-Lempel algorithm for data compression. In cryptography it was applied by Leung and Tavares [Leun 85] for testing block ciphers. In his PhD. thesis [Wan 88] M. Wang mentions some of the properties of the Ziv-Lempel complexity which are also part of the work in this paper but he does not prove this properties.

D.W. Davies (Ed.): Advances in Cryptology - EUROCRYPT '91, LNCS 547, pp. 114-126, 1991.
© Springer-Verlag Berlin Heidelberg 1991

In our paper we will consider minimal and maximal values of the Ziv-Lempel complexity. After that we will have a closer look at the Ziv-Lempel complexity for periodic pseudorandom number sequences. Particularly, we will see that the Ziv-Lempel complexity of these sequences depends on the start position s_j where the computation of the complexity has started and that the Ziv-Lempel complexity for such a sequence has a constant value after maximal 2p - 1 positions of the sequence have been considered (p denotes the period of the sequence). After that we will consider the Ziv-Lempel complexity for arbitrary pseudorandom number sequences. Finally, the Ziv-Lempel complexity will be compared with other well-known cryptographic complexity measures such as the linear complexity or the maximal-order complexity. Here, we will mainly consider the question whether it is necessary to examine pseudorandom number sequences using the Ziv-Lempel complexity or whether it is enough to use only the linear complexity.

2 Definition and Computation of the Ziv-Lempel complexity

As we mentioned above Ziv and Lempel introduced a complexity measure for finite sequences in 1976. Intuitively, this complexity is a measure of the rate at which new patterns emerge as we move along the sequence. We refer the reader to [Lem 76] for the formal description. In this paper we will only provide a short description as it is given in [Ziv 78].

Let $S = s_1 \ldots s_n$ be a sequence of length n then the following rules can be used to obtain the Ziv-Lempel complexity of the sequence S:

1) A slash is inserted following s_1
2) Assume that the i-th slash comes after the letter s_{k_i}, $1 \leq k_i \leq n - 1$. The next slash will be inserted after the letter s_{k_i+1} where $k_{i+1} = k_i + L_i + 1 \leq n$ and L_i is the maximal length of a substring $s_{k_i+1} \ldots s_{k_i+L_i}$ such that there exists an integer p_i (where $1 \leq p_i \leq k_i$) for which $s_{p_i} \ldots s_{p_i+L_i-1} = s_{k_i+1} \ldots s_{k_i+L_i}$.

If s_n is followed by a slash the Ziv-Lempel complexity is equal to the number of slashes, otherwise the Ziv-Lempel complexity is equal to the number of slashes plus one.

To illustrate the mechanism for the computation of the Ziv-Lempel complexity we will use the following example: let X = 1001101110000111 be a binary sequence then we can insert slashes into the sequence X using the two rules and we will obtain the sequence X with the following partition X = 1|0|01|101|1100|00111|. X is now devided into six patterns and therefore the Ziv-Lempel complexity C for the sequence X is equal to 6.

In the next step we will present an algorithm for the computation of the Ziv-Lempel complexity. To compute the Ziv-Lempel complexity we have to define an algorithm for the execution of rule 2. Rule 2 can be executed using the following two steps:

a) Initialization: Let $n + 1$ be the sequence position where the computation of a new pattern starts, let J contain all positions j for which $s_j = s_{n+1}$ $(1 \leq j \leq n)$ and let $l = 2$.

b) Repeat the following step until J is empty: delete all $j \in J$ for which $s_{j+l-1} \neq s_{n+l}$ and increase l by 1 if J is not empty. Otherwise l is the length of the new pattern which is defined by $s_{n+1}...s_{n+l}$.

These two steps enable us to compute the patterns of a sequence S and therefore also the Ziv-Lempel complexity of the sequence.

To improve the computational complexity of the computation of the Ziv-Lempel complexity pattern recognition algorithms such as the algorithm of Blumer, Blumer, Ehrenfeucht, Haussler and McConnel [Blu 83] can be used.

3 Minimal and Maximal value of the Ziv-Lempel complexity

In this section we consider a binary sequence $S = s_1..s_n$ of length n which consists of k ones and n - k zeros $(0 \leq k \leq n)$. Our aim is to examine the minimal resp. maximal value of the Ziv-Lempel complexity of this sequence.

For the minimal value of the Ziv-Lempel complexity we obtain the following Theorem.

Theorem 1: The minimal value of the Ziv-Lempel complexity C_{min} of a sequence $S = s_1...s_n$ is

$$C_{min} = \begin{cases} 2 & \text{if } k = 0 \text{ or } k = 1 \text{ or } k = n \text{ or } k = n-1 \\ 3 & \text{in all the other cases} \end{cases}$$

Proof: Computing the Ziv-Lempel complexity for a sequence S we have to use s_1 as the first pattern. To get now a minimal Ziv-Lempel complexity $C_{min} = 2$ we have to use $s_2...s_n$ as a second pattern. This is only possible if the positions s_j, $2 \leq j \leq n-1$, have the same value as position s_1. If the value of position of s_n is different to the value of position s_1 we get an unique pattern $s_2...s_n$, otherwise the pattern will not be unique. Depending on the way how s_1 was chosen there will be $k = 0$, $k = 1$, $k = n-1$ or $k = n$ ones in the sequence.

For any other choice of k at least three patterns have to be used to get the Ziv-Lempel complexity of the sequence and therefore the possible minimal value of the Ziv-Lempel complexity C_{min} is 3. To obtain now a sequence with minimal Ziv-Lempel complexity the sequence can be constructed in the following way: Under the assumption that s_1 is equal to 1 the first pattern p_1 is again equal to s_1, the second pattern p_2 is equal to a (k-1)-times repetition of p_1 followed by the inverse value of p_1 and the third pattern p_3 consists only of a (n - k)-times repetition of the inverse value of p_1. If s_1 is equal to 0 the same procedure can be done but p_2 contains then (n - k - 1) repetitions of p_1 and p_3 (k -1). This is only one principle to construct such sequences. It is also possible to have one 1 as value of the final position of pattern p_3 and therefore to reduce the numbers of ones in pattern p_2 by 1. The same principle can also be applied to the zeroes, if the sequence starts with a zero.∎

For a further construction principle of sequences with $C_{min} = 3$ let $l_1 = 1$ be the length of pattern p_1, l_2 the length of pattern p_2 and $l_3 = n - l_1 - l_2$ the length of pattern p_3 and $k < n/2$. Furthermore let $p_1 = 0$, $p_2 = 0...01$ and p_3 be a repetition of the final l positions of p_2 for (m -1)-times followed by zeros or for (m - 2)-times where the last position has the value 1, then we obtain additional sequences with $C_{min} = 3$ if for l_1, l_2 and l_3 the following equation holds:

$$(m-2)\cdot l + 1 \leq n - l_1 - l_2 \leq m \cdot l, \qquad l \leq l_2$$

The following sequences are examples which are constructed using these three principles:

S = 11.....10 which has Ziv-Lempel complexity C = 2.
S = 11...100...0 which has Ziv-Lempel complexity C = 3.
S = 00...011..10 which has the Ziv-Lempel complexity C = 3.
S = 000...001001....00100 which has the Ziv-Lempel complexity C = 3.

After having considered the minimal value of the Ziv-Lempel complexity for a sequence S, in a next step we will now examine its maximal value.

Whereas it is possible to make some assumptions about the minimal value of the Ziv-Lempel complexity for any k, $1 \leq k \leq n$, this is not possible in the case of the maximal value of the Ziv-Lempel complexity. Here we can only give the maximal value of the Ziv-Lempel complexity for some certain values of k.

Lemma 1: Let n be the sequence length and

$$n = \sum_{\substack{i=0 \\ i \text{ even}}}^{k} l_i + l_{i+1} + r \qquad (1)$$

with $l_0 = l_1 = 1$, $l_i = l_{i-2} + l_{i-1}$ and $l_{i+1} = l_{i-2} + l_{i-1} + 1$ for $2 \leq i \leq k$ and $0 \leq r < l_{k+2} + l_{k+3}$ then the maximal value of the Ziv-Lempel complexity C_{max} is

$$C_{max} = \begin{cases} k & \text{for } r = 0 \\ k+1 & \text{for } 1 \leq r \leq l_{k+2} \\ k+2 & \text{for } l_{k+2} \leq r \leq l_{k+2} + l_{k+3} \end{cases}$$

if the number of ones or zeros in the sequence is equal to k/2.

Proof: Using formula (1) we get a construction criterion for a sequence of length n which contains k/2 ones and has Ziv-Lempel complexity C_{max}. If the patterns p_i (i even) of length l_i only consist of zeros and the patterns p_{i+1} of length l_{i+1} have the form 00..001, then each of these patterns defines a unique pattern as it is necessary to get the Ziv-Lempel complexity and the Ziv-Lempel complexity of the sequence is equal to C_{max}.

It is not possible to obtain a Ziv-Lempel complexity C which is greater than C_{max} under the assumption that the sequence length is equal to n and that the number of ones which the sequence contains is equal to k/2; because the number of patterns used to compute the Ziv-Lempel complexity will be maximal if each one appears in a different pattern and if each of the patterns containing a single one is followed by a pattern containing no one. After the rules for the construction of the patterns a pattern containing no one has to be followed by a pattern containing a single one and therefore it is not possible to have more than k patterns without a single one. ∎

If the number of ones in the sequence is less than k/2 the maximal Ziv-Lempel complexity can be obtained using the same construction method.

4 Ziv-Lempel complexity for periodic sequences

In this section we consider the Ziv-Lempel complexity for periodic sequences $s_1...s_{p-1}s_ps_{p+1}...$ where p is the period of the sequence and $s_{p+i} = s_i$ for $0 < i \leq p$. This type of sequence is often generated by pseudorandom number generators but can also be obtained if we consider message files which consist of a repetition of one message for a certain number of times. In our theorems we will obtain the maximal length l of a periodic sequence which is needed to compute the Ziv-Lempel complexity for the sequence of any length greater than l.

In the following we will start with the consideration of a special case of sequences and then continue our examination with the general case of periodic sequences. We assume for our first theorem that the sequence $s_1....s_p$ has Ziv-Lempel complexity C and the sequence $s_1....s_{p+1}$ has Ziv-Lempel complexity C + 1. Using this condition we see that one period of the sequence

finishes with a complete pattern as it is defined by the Ziv-Lempel complexity. Then we obtain the following theorem.

Lemma 2: If S is a periodic sequence where $s_1...s_p$ has the Ziv-Lempel complexity C and $s_1...s_{p+1}$ has the Ziv-Lempel complexity C + 1 then for $l \geq q = p + 1$, the sequence $s_1...s_l$ will have the Ziv-Lempel complexity C + 1.

Proof: Because of the period of the sequence s_{p+i} has the same value as s_i for $i \geq 1$ and because of the start of a new pattern at position p + 1 the pattern given by $s_{p+1}...s_{p+j}$ can allways be found as $s_1...s_j$ for $j \geq 1$ in the sequence. Therefore for any $j \geq 1$ the increase of j by one does not define a new pattern and therfore the Ziv-Lempel complexity can never be greater than C + 1 for $j \geq 1$.∎

Now we will consider the more general case where it is only known that the sequence S has a period p.

Lemma 3: If S is a sequence with period p, Ziv-Lempel complexity C for $s_1...s_{p-k}$ and Ziv-Lempel complexity C + 1 for $s_1...s_{p+1}$ then there exists a q such that the sequence $s_1...s_j$ has the Ziv-Lempel complexity C + 1 for $p - k < j \leq q$ and the Ziv-Lempel complexity C + 2 for $j > q$. $(k > 0)$

Proof: Because of the period of the sequence the pattern starting at position p - k + 1 cannot be of infinite length. Therefore a position q = p + l must exist in such a way that $s_{p-k+1}...s_{p+l}$ defines a unique pattern. All sequences $s_1...s_j$ have the same Ziv-Lempel complexity C + 1 for $p - k + 1 \leq j \leq p + l = q$.

Because of the period of the sequence the pattern starting at position p + l + 1 can also be found with starting position l + 1. Because of the periodicity of the sequences this is true for every pattern length $j \geq 1$. Therefore all these sequences have the same Ziv-Lempel complexity C + 2.∎

In our next two lemmas we will determine a maximum for the pattern length of the pattern $s_{p-k+1}...s_{p+l}$ and using this maximum we can provide an upper bound of the sequence length which has to be examined in the case of periodic sequences to obtain the Ziv-Lempel complexity of the sequence.

Lemma 4: Using the assumptions of lemma 2 the maximum length m for the pattern $s_{p-k+1}...s_{p+l}$ is p - k.

Proof: The following equitation has to be true to obtain a pattern

$$s_{p-k+l} \neq s_{j+l}$$

and

$$s_{p-k+m} = s_{j+m} \quad \text{for } 0 < m < l$$

where j is called the starting point of the existing pattern $s_j...s_l$. Now let us assume $l > p - k$ then we obtain the following additional equation

$$s_{j+l} = s_{j+l-p+k}$$

which itself provides us a new equation

$$s_{j+l-p+k} = s_{p+j+l-p+k} = s_{j+l+k}$$

because of the period p of the sequence. Starting with s_{j+l+k} we obtain

$$s_{j+l+k} = s_{j+l-p+k+k} = s_{p+j+l-p+2k} = s_{2j+2k+l}$$

Continuing this chain we obtain

$$s_{j+k+l} = s_{2j+2k+l} = \ldots = s_{xj+xk+l} \qquad \text{for } x > 0$$

Now we have to consider

$$s_{p-k+l} \neq s_{xj+xk+l}$$

In a first step we will show that for fixed $l > p - k$ and certain choices of j and k the two values are equal and therefore it is not possible to have a pattern length $l > p - k$. For simplification let us only write i when we actually mean s_i.

$$p - k + l = xj + xk + l$$

If for some $y > 0$, $\underline{x} > 0$:

$$yp - k + l = \underline{x}j + \underline{x}k + l$$

because we have a periodic function. We obtain now

$$yp = \underline{x}j + (\underline{x} + 1)k \leftrightarrow \underline{x} = (yp - k)/(j * k) \qquad (1.1)$$

So if there exists y for p and k such that the value of the fraction (1.1) is an integer, $p - k + l$ can never be chosen different from $j + l$, and therefore it is not possible to have a pattern length greater than $p - k$.
Let us now consider the case in which (1.1) has no integer solution. Because of the definition of the Ziv-Lempel complexity and the period of the sequence we get the following additional equitations:

$$s_1 = s_{p+1} = s_{j+k+1} = s_{p+j+k+1} = s_{2j+2k+1} = \ldots = s_{xj+xk+1}$$

$$s_2 = s_{p+2} = s_{j+k+2} = \ldots = s_{xj+xk+2}$$

Continuation leads to

$$s_{k+j} = s_{p+k+j} = s_{j+j+k+k} = \ldots = s_{xj+xk}$$

$$s_{k+j+1} = s_{p+k+j+1} = s_{j+j+k+k+1} = \ldots = s_{xj+xk+1}$$

therefore we would only obtain a sequence with period $j + k$ instead of period p. It is not possible to have a patternlength $l > p - k$ and period p.∎

Lemma 5: $q \leqq 2p - 2$ under the assumption made in lemma 3.

Proof: Follows from lemma 3 using $k = 1$ which is the maximum pattern length l.∎

Now we examine at which position j the pattern $s_{p-k+1}...s_{p+l}$ starts in the part $s_1...s_{p+k}$ of the sequence. We see, the number of j's will be limited and will depend on the value of k and the length of the pattern l.

Lemma 6: Using the assumption made in lemma 3 for fixed k and $l \leq p - k$, j has to be chosen in such a way that $j + l \leq p - k$.

Proof: The same proof construction as in lemma 3 can be used if the assumption $l > p - k$ is replaced by the assumption $j + l > p - k$.∎

Using lemma 5 we get the number of sequences with period p and $q = 2p - 2$. For $l = p - k = p - 1$, j has to be equal to 0 and therefore we can obtain exactly 2 sequences for this condition:

S1 = 0|0...01|00...0|10 where the 1 is at position p - 1

S2 = 1|1...10|11...1|01 where the 0 is at position p - 1

S2 is the complement sequence to S1. For both sequences C is equal to 3.

Theorem 2: In a sequence with period p the Ziv-Lempel complexity C has a constant value after maximal 2p - 1 positions of the sequence have been considered.

Proof: Follows immediately from lemma 1 - 5.∎

Theorem 2 shows that it is sufficient to use 2p -1 positions of a periodic sequence for the computation of the Ziv-Lempel complexity.

We will now undertake a closer examination of the starting points for the computation of the Ziv-Lempel complexity. We will see that the Ziv-Lempel complexity for a periodic sequence depends on the starting point and that therefore we can obtain different values for the Ziv-Lempel complexity for different starting points.

Lemma 7: Let S be a periodic sequence with period p, let C_1 be the Ziv-Lempel complexity of the sequence computed with start position s_i and C_2 be the Ziv-Lempel complexity of the sequence computed with start position s_{i+1} then C_1 and C_2 can be different.

Proof: In our proof we show only one possible situation where C_1 and C_2 will be different. Let S_1 be the periodic sequence used to compute C1 with $s_1 \neq s_2$ and $s_2 = s_3$, then we obtain a pattern $p_1 = s_1$, a pattern $p_2 = s_2$ and further patterns p_i for which we assume that $p_i = s_i...s_{i+k}$ and $s_i...s_{i+k-1}$ is equal $s_j...s_{j+k-1}$ for $j > 1$ ($i > 2$, $k > 0$) or in words no pattern should be built using a already existing pattern starting at position 1 of the sequence S_1. C_2 will now be computed using starting position s_2.Therefore we loose the pattern $p_1 = s_1$. Under the assumption that for the last limited pattern of S_1 $p_{C_1-1} = s_{p-k}...s_{p+j}$ a similar pattern can be found in S_2 which has the form $s_{p-k-1}...s_p s_{p+1}...s_{p+j}$

$(k, j, i > 0)$. We actually reduce the number of patterns in S_2 by one because we lost the inital pattern p_1. Therefore the Ziv-Lempel complexity C_2 is equal to $C_1 - 1$ and therefore C_2 is not equal to C_1.∎

This result shows us a great difference between the behaviour of the Ziv-Lempel complexity and of other complexity measures such as the linear complexity for sequences or the maximum order complexity.

Now we will look at the Ziv-Lempel complexity values C_i computed for a periodic sequence where C_i is computed using start position s_i $(1 \leqq i < p)$. We see that the values C_i are distributed around the average value $\underline{C} = 1/n\,(C_1 + \ldots + C_{p-1})$. Examing the value C_i we get with the highest probability the value $\underline{C}$ and with exponentially decreasing probabilities the complexity values $\underline{C} - j$ and $\underline{C} + j$ $(j > 0)$. We also see that $\underline{C} - j$ is bounded by a value C_{min} called minimal Ziv-Lempel complexity of the periodic sequence and $\underline{C} + j$ by a value C_{max} called maximal Ziv-Lempel complexity of the periodic sequence. Having $\underline{C}$ fixed C_{min} and C_{max} will also be fixed because of the structure of the sequence defined by C.

5 Ziv-Lempel complexity for binary pseudorandom number sequences

The Ziv-Lempel complexity can be used to examine pseudorandom number sequences. In the following let us assume $p(s_j = 1) = p(s_j = 0) = 0.5$.

In a first step we will examine the average length of a pattern which starts at position s_{n+1}.

Lemma 8: Under the assumption that each position of the sequence can be computed independently of any previous position the average length of a pattern starting at position s_{n+1} is $p_a = \lfloor \log_2 n \rfloor + k$ where $\lfloor x \rfloor$ denotes the integer part of x and k is equal to $k = ((1 - c_0 \cdot 1 + c_0 \cdot 2^{-1} \cdot 2 + c_0 \cdot 2^{-2} \cdot 3 + \ldots) / (1 - c_0 + c_0 \cdot 2^{-1} + c_0 \cdot 2^{-2} + \ldots)$ and c_0 can be computed using the following formula: $c_0 = n$; for $j = 1$ to $\lfloor \log 2n \rfloor + 1$ do $c_0 = c_0 / 2$.

Proof: The average length of a pattern is equal to the number of steps which are needed on average until the set J is empty. Under the assumption that the probability $p(s_j = 1) = p(s_j = 0) = 0.5$ we obtain on average n/2 values $j \in J$ in step a) of the algorithm given in section 2. As long as there are 2 or more elementes in J in step b) of this algorithm the number of j's $\in J$ will be halved on average if l is increased by 1. This is the case for $\lfloor \log_2 n \rfloor$ steps. If J contains for the first time less than two positions a new pattern exists in a certain number of cases. The number is $1 - c_0$ where c_0 is computed by the formula given above. After that in each round r we will obtain new patterns with length $\lfloor \log 2n \rfloor + r + 1$ in $c_0 \cdot 2^{-r}$ cases where $0 < r$. On average k rounds are

needed until a new pattern is obtained. Therefore the pattern length of the new pattern starting at position $n+1$ will have on average the length $\lfloor \log_2 n \rfloor + k$.∎

Remark: $2 \leq k < 3$

For $n = 2^{-m}$ $(0 < m)$ we obtain $c_0 = 0.5$ and therefore $(\sum_{k=1}^{\infty} 2^{-k} \cdot k) / \sum_{k=1}^{\infty} 2^{-k}$

If $c_0 \to 0$ we obtain $\lim ((1-c_0) \cdot 1 + c_0 \cdot 2^{-1} \cdot 2 + c_0 \cdot 2^{-2} \cdot 3 + \ldots) / (1 - c_0 + c_0 \cdot 2^{-1} + c_0 \cdot 2^{-2} + \ldots) = (\sum_{k=1}^{\infty} 2^{-k} \cdot (k+1)) / \sum_{k=1}^{\infty} 2^{-k} = 3$

Having an estimation of the average pattern length for a pattern starting at position j, $1 \leq j \leq n$ we can now give an estimation of the average Ziv-Lempel complexity Ca for a sequence of length n under the assumption that each pattern of the sequence is computed independently of the former patterns. The estimation can be given using the following induction:

1) For j := 1: $C_a := 1$ and $j_0 := 1$
2) For j := 2...n: if $j = j_0 + \lfloor \log_2 j_0 \rfloor + k$ then $C_a := C_a + 1$ and $j_0 := j$

Both measures the Ziv-Lempel complexity C_a and the pattern length pa can be used to examine pseudorandom number sequences. We would expect that a good pseudorandom number sequence has a Ziv-Lempel complexity which is close to the value C_a and that each pattern of the sequence has a length p_i which is close to the corresponding p_{a_i}. If the Ziv-Lempel complexity of the sequence is much smaller than the Ziv-Lempel complexity C_a then the sequence will have at least one multiple repetition of a sequence pattern and therefore there exists at least one pattern length p_i such that p_i is much greater than p_{a_i}. The pattern i will be a repetition of already existing patterns. This should be avoided for two reasons:

1) Often the pattern i might repeat a pattern which does not include the same number of ones and zeros for several times and therefore the ones and zeros might locally not be randomly distributed with $p(s_j = 1) = p(s_j = 0) = 0.5$

2) The multiple repetition of a pattern can also be a disadvantage if the pseudorandom number sequence is used in a stream cipher. If not only the pseudorandom number sequence has a multiple repetition of patterns but also the plaintext, the corresponding ciphertext might also have a multiple repetition of patterns.

Therefore the pattern length p_i and the Ziv-Lempel complexity can be used to indicate pseudorandom number sequences which are not desirable.

6 Ziv-Lempel complexity and other cryptographic complexity measures

In this paper we only want to compare the Ziv-Lempel complexity with the linear complexity. A comparision of the Ziv-Lempel complexity with the maximal-order complexity would have a similar effect.

Looking only at the linear complexity we see the sequence can have a great linear complexity which is suitable but on the other side the same sequence can have a small value for the Ziv-Lempel complexity which is not desirable. An example would be the sequence S = 00...001. Therefore the value for the linear complexity alone does not say much about the quality of the sequence. But considering also the Ziv-Lempel complexity for this kind of sequences the value for the Ziv-Lempel complexity will be close to the minimal Ziv-Lempel complexity and therefore the sequence would not be suitable.

If the linear complexity profile is considered instead of the linear complexity we obtain much better statements about the quality of the sequence and the Ziv-Lempel complexity can be seen as an additional measure providing us with some more information about repetitions of patterns in the sequence. Particularly, the Ziv-Lempel complexity can be used to detect repetitions of patterns.

Let us now consider the case of R repetitions of a pattern in the sequence $S = s_1...s_k$ and let us define the complexity length for the linear complexity. If s_i is the first position in the sequence with linear complexity C_i and if s_j is the first position with linear complexity $C_i + 1$ then $j - i$ will be considered as the complexity length.

Now we consider a sequence $S = s_1...s_k$ which has the following Ziv-Lempel complexity $S = s_1|s_2...|s_i...s_k|$ where $s_i...s_k$ consists of $R > 1$ repetitions of a pattern $s_j...s_{i-1}$ where $1 \leqq j \leqq i - 1$. Let us consider the linear complexity and therefore the complexity length of this sequence. If $s_i...s_k$ is equal to $s_1...s_{i-1}\ s_1...s_{i-1}...s_1...s_{i-1}s_k$ then there exists an $\underline{i}$ where $i < \underline{i} < 2i$ according to the well-known results for the linear complexity in periodic sequences such that the linear complexity C_l is the same for $s_1...s_j$, $\underline{i} \leqq j \leqq k$. The complexity length $k - \underline{i}$ will then indicate that there is an unregularity in the sequence because $k - \underline{i}$ will be large in comparision to the other complexity lengths in the squence.

If $s_i...s_k$ is equal to $s_j...s_{i-1}s_j...s_{i-1}...s_j...s_{i-1}...s_k$ then the theorems about periodic sequences and the linear complexity canot be used and the linear complexity for the sequence $s_1...s_k$ will normally increase in such a form that there is no indication that the pattern $s_j...s_{i-1}$ is repeated for R times.

Considering these two cases we see the linear complexity and the complexity length alone are not enough to discover all kinds of repetitions of patterns which are possible in a sequence. Therefore we need an additional complexity measure which enables us to detect other kind of repetitions of patterns which do not influence the linear complexity. The Ziv-Lempel complexity together with the pattern length can be used as such a complexity measure.

7 Conclusions

In cryptanalysis it should be possible to use a complexity measure for several different tasks such as the examination of pseudorandom number sequences or the identification of a period in a sequence. Therefore it is necessary to determine requirements which help to identify complexity measures which are good for all these different tasks. In the following we will present four points which we think a complexity measure should fullfil:

1) Indication and identification of a period in a sequence: The complexity measure should be able to detect a period in a sequence and to identify its starting point and its length.

2) Independence of the complexity value in a periodic sequence from the position where the computation had started: If in a periodic sequence the complexity value is computed using different start positions, the result should be the same after a certain number of positions had been considered.

3) Examination of pseudorandom number sequences: It should be possible to use the complexity measure to examine pseudorandom number sequences and to compare the results against results obtained from random sequences.

4) Computation of the sequence: Knowing the complexity of the sequence it should be possible in an easy and efficient way to compute the sequence.

If we now look at the Ziv-Lempel complexity and compare our results with this four points we see that the Ziv-Lempel complexity fullfils some of these points but that there are still some points with great difficulties. For example, the Ziv-Lempel complexity does not fullfil points 2) and 4). However, the Ziv-Lempel complexity can be used to indicate a period in a sequence and it can be used to examine pseudorandom number sequences. Here there are still some problems to be solved because it is far from clear how the distribution of Ziv-Lempel complexity behaves for a certain sequence length.Therefore, additional examinations of the Ziv-Lempel complexity measure in cryptography are necessary.

Nevertheless, the Ziv-Lempel complexity is usefull in cryptography because as it is shown in section 6, it can detect weaknesses in a sequence which can not be detected by the linear complexity.

8 References

[Blu 83] A. Blumer, J. Blumer, A. Ehrenfeucht, D. Haussler, R. McConnel: Linear Size Finite Automata for the Set of all Subwords of a Word, An Outline of Results, Bul. Eur. Assoc. Theor. Comp. Sci., No. 21, 1983, pp. 12-20

[Jans 89] C. Jansen: Investigations On Nonlinear Streamcipher Systems: Construction and Evaluation Methods, PhD. Thesis

[Lemp 76] A. Lempel, J. Ziv: On the Complexity of Finite Sequences, IEEE Transaction on Information Theory, IT-22, No.1

[Leun 85] A. Leung, S. Tavares: Sequence Complexity as aTest for Cryptographic Systems, Advances in Cryptology - Crypto 84, lecture Notes in Computer Science, springer Verlag, Heidelberg, 1985

[Ruep 86] R. Rueppel: Analysis and Design of Stream Ciphers, Springer Verlag, Berlin, 1986

[Wan 88] M. Wang: Cryptographic Aspects of Sequence Complexity Measures, PhD. Thesis, ETH Zürich, 1988

[Ziv 78] J. Ziv: Coding Theorems for Individual Sequences, IEEE Transaaction on Information Theory, IT-24, No. 4