ON THE LINEAR COMPLEXITY OF COMBINED SHIFT REGISTER SEQUENCES.

Lennart Brynielsson
Fst/TSA
Box 80001
S-104 50 STOCKHOLM SWEDEN

Many proposed keystream generators consist of a number of binary maximum length shift registers combined by a nonlinear binary function. The registers guarantee a long period and the nonlinear function destroys the linearity i.e. it gives the output sequence a large linear complexity <1>, (linear equivalent <2>). In order to avoid correlation attacks the function should also be correlation immune <3> i.e. the output sequence should be statistically independent of the various inputs. There is however a trade off between the linear complexity and the order of correlation immunity, since it is not easy to achieve both properties. The reason for this is that in the binary field GF(2) there are too few functions. As an example the only correlation immune function of two variables is linear.

In the field $GF(2^e)$ the situation is different. For instance, the polynomial function $x+y+3xy+2(x^2y+xy^2)+x^2y^2$ in GF(4) is both nonlinear and correlation immune. In order to valuate such a function one must be able to calculate its linear complexity. That is the purpose of this paper. We shall show the following result stated here for two variables.

THEOREM: Let x and y be two sequences in $GF(2^e)$ given by maximum length shift registers of lengths m and n which are relatively prime and greater than three. If they are combined by means of a polynomial function the linear complexity L of the resulting sequence is given by

F. Pichler (Ed.): Advances in Cryptology - EUROCRYPT '85, LNCS 219, pp. 156-160, 1986.

$$L\left(\sum A_{ij} x^i y^j\right) = \sum_{A_{ij}\neq 0} m^{\|i\|} n^{\|j\|} \qquad A_{ij} \in GF(2^e)$$

where $\|i\|$ is the number of ones in the binary representation of i. This result is general since all functions in a finite field are polynomial functions <1>. We shall also sketch a generalization to $GF(p^e)$ for p>2.

Example

If the correlation immune polynomial mentioned above is used to combine two registers in GF(4) of length 4 and 5 then the linear complexity is 4+5+20+20+20+20. In fact the polynomial implements the function "x plus y mod 4". A more striking example is obtained if "x plus y mod 16" in GF(16) is written as a polynomial. It turns out to be correlation immune and it contains many nonlinear terms. If GF(16) is implemented as $GF(2)(t)/(t^4+t+1)$ and two registers of length 17 and 19 are combined, the linear complexity is 1670090.

Preliminaries

We shall use the following results which have been proved more generally by among others Selmer <4>, Herlestam <1>, Zierler and Mills <5>.

Lemma: Consider two sequences from two linear feedback registers whose feedback polynomials have simple roots a_i and b_j which are all different. The sum of the sequences will have a feedback polynomial the roots of which constitute the union $\{a_i, b_j\}$. Moreover, if all root pair products a_ib_j are different then the product sequence will have a feedback polynomial with roots a_ib_j. This can be seen from the fact that the output n-th term from such sequences can be written as a linear combination of the n-th powers of the roots of the feedback polynomial <6>.

Proof of the theorem

We work in $K=GF(2^e)$. Let the sequence $x=(x_i)$ be generated by a linear shift register with maximum length feedback polynomial f of degree m. This implies that if a denotes a root of f then the extension field $K(a) = GF(2^{em})$ consists of the elements

$0, a, a^2, \ldots, a^{2^{em}-1} = 1$. The polynomial f has the following roots

$$a \quad a^{2^e} \quad a^{2^{2e}} \quad \ldots\ldots \quad a^{2^{(m-1)e}}$$

Now consider the sequence $x^2 = (x_i^{\,2})$. Squaring is an automorphism in fields of characteristic two and therefore this sequence must be obtained if you square the coefficients of f. The roots will also be squared. Repeating this procedure gives that the m roots of the polynomial which generates x^{2^k} are

$$a^{2^k} \quad a^{2^{e+k}} \quad \ldots\ldots \quad a^{2^{(m-1)e+k}} \qquad k=0,\ 1,\ \ldots,\ e-1$$

Thus the roots are of the form a raised to different powers of two: 2^i, $i=0,1,\ldots me-1$. Consequently, if you multiply a number of different roots then their exponents will add and it is possible to deduce from the resulting exponent which roots that were used as factors. Different factors give different products.

Now consider the sequence x^n where $0 < n < 2^e$. The exponent n can be written as a sum of powers of two and the sequence can be looked upon as a product of the corresponding sequences x^{2^k} which have been described above. The root products are different and we can use the lemma of Selmer <4>. We have now proved:

Theorem: Let the sequence x in $GF(2^e)$ be given by a maximum length polynomial of degree m. Then a polynomial sequence has the linear complexity

$$L\ (\sum_i A_i x^i) = \sum_{A_i \neq 0} m^{\|i\|} \qquad A_i \in GF(2^e)$$

We note also that the root products do not belong to K since K consists of the elements $0,\ a^r,\ a^{2r}, \ldots\ldots\ a^{(2^e-1)r}$ where $r=(2^{em}-1)/(2^e-1)$ and these powers are obtained when all roots belonging to one x^n - sequence are multiplied together.

Consider now two sequences x and y over $K=GF(2^e)$ given by maximal length polynomials with degrees m and n which are supposed to be relatively prime. The common splitting field of those polynomials is $GF(2^{emn})$. We denote the primitive roots a and b. Both K(a) and K(b) are subfields and $K(a) \cap K(b) = K$ since the intersection consists of those elements in $GF(2^{emn})$ which remain fixed under the automorphisms $t \to t^{2^{em}}$ and $t \to t^{2^{en}}$. Therefore they are also fixed under the automorphism $t \to t^{2^{e\gcd(m,n)}} = t^{2^{e}}$ which implies that they belong to $K=GF(2^{e})$.

When a term $x^k y^l$ is formed we obtain root products of the type $a^i b^j$ where a^i originates from x^k and similarly b^j from y^l. Again we must show that different factors give rise to different products. If $a^i b^j = a^{i'} b^{j'}$ then it follows that $a^{(i-i')} \in K$. That this is impossible when $m \geq 4$ can be seen as follows. Arrange for an element in $K(a) = GF(2^{em})$ the binary representation of the exponent of a in a exm-matrix. Then for elements of K each row will consist entirely of either zeros or ones whereas for elements which are root products there will be at most a single one in each row. For $m \geq 4$ it is impossible for the sum of a "K type" and a "root product type" exponent to yield another "root product type"; there will be too many ones left.

The case $GF(p^e)$ when $p>2$

Similar results are also valid when the characteristic is greater than two i.e. when $K=GF(p^e)$. The difference in the deduction when $p>2$ concerns the roots corresponding to x^k for $1<k<p$. Herlestam has shown <7> that all $C(m+k-1,k)$ possible root products are present. By means of automorphisms it can be shown that the polynomial corresponding to x^{kp^j} has the same number of roots, all of the form a^i where i in the p-ary number system has only one single nonzero digit. The linear complexity of a power x^n, $0<n<p^e$, can now be written:

$$L\ (x^n) = \prod_{i=0}^{e-1} \binom{m+n_i-1}{n_i}$$

where $n_0, n_1, \dots n_{e-1}$ denote the p-ary digits. This expression is the generalisation of $m^{\|i\|}$ when $p>2$. The condition greater than three should be replaced by greater than p+1. In all other respects the proofs and theorems are similar.

References

<1> T. Herlestam, "On the Complexity of Functions of Linear Shift Register Sequences", IEEE 1982, Les Arcs, France.

<2> E.J. Groth, "Generation of Binary Sequences with Controllable Complexity", IEEE Trans. on Inf. Th. It-17 1971.

<3>. T. Siegenthaler, "Correlation Immunity of Nonlinear Combining Functions for Cryptographic Applications." IEEE Trans. on Inf. Th. It-30 1984.

<4> E.S. Selmer, "Linear Recurrence Relations over Finite Fields", Dept of Math., Univ. of Bergen, Norway, 1966.

<5> N. Zierler and W.H. Mills, "Products of Linear Recurring Sequences", J. Algebra, 27, 1973.

<6> T. Beth, "Stream Ciphers", Proceedings of Secure Digit Comm. C.I.S.M. Udine 1982.

<7> T. Herlestam, private communication, to be published.