On the use of expansion series for stream ciphers

Claus Diem

April 5, 2012

Abstract

From power series expansions of functions on curves over finite fields, one can obtain sequences with perfect or almost perfect linear complexity profile. It has been suggested by various authors to use such sequences as key streams for stream ciphers. In this work, we show how long parts of such sequences can be computed efficiently from short ones. Such sequences should therefore considered to be cryptographically weak. Our attack leads in a natural way to a new measure of the complexity of sequences which we call expansion complexity.

1 Introduction

It is well known that linearly recurrent sequences, that is, sequences generated from linear feedback shift registers (LFSR), are cryptographically weak. This observation leads to the following well-established definitions; cf. [18], [16] and other works on linearly recurrent sequences.

We set $\mathbb{N} := \{1, 2, \ldots\}$. Let q be a prime power. By a *sequence* over $\mathbb{F}_q$ we mean a map from a subset of the form $\{1, \ldots, m\}$ or from $\mathbb{N}$ to $\mathbb{F}_q$. For a finite sequence $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ over $\mathbb{F}_q$ one defines the *linear complexity*, $L_{\mathbf{a}}$, as the least ℓ such that $\mathbf{a}$ is generated by a linear recurrence relation of order ℓ. Now, for a finite or infinite sequence $\mathbf{a} = (a_1, a_2, \ldots)$ of length m over $\mathbb{F}_q$ and for $n \leq m$, one defines $L_{\mathbf{a}}(n)$ as the linear complexity of the finite subsequence consisting of the first n terms of $\mathbf{a}$, and one defines the *linear complexity profile* as $(L_{\mathbf{a}}(n))_{n=1}^{m}$.

An infinite sequence $\mathbf{a} = (a_i)_{i \in \mathbb{N}}$ over a finite field $\mathbb{F}_q$ is said to have *perfect linear complexity profile* if $|2L_{\mathbf{a}}(n) - n| \leq 1$ for all n and *d-almost perfect linear complexity profile* if $|2L_{\mathbf{a}}(n) - n| \leq d$ for all n.

In [20] a general construction of sequences with almost perfect complexity profile was given. The construction is based on function expansion into expansion series and will be recalled below. The motivation stated in [20] to consider this construction is the generation of key streams for stream

ciphers. Also in [16] the consideration of the construction is motivated by applications to stream ciphers.

In this work, we show that the coefficients of the sequences constructed via the method in [20] can be efficiently computed from relatively short subsequences. Therefore, the sequences should be considered as cryptographically weak.

The proposed attack leads to a new notion of complexity which we call *expansion complexity*. The expansion complexity of a sequence is always at most the linear complexity and captures the immunity against our attack.

Additionally, we show how one can apply well known results on continued fraction expansion in order to obtain further inside into the series suggested in [20]. In particular, we refute the conjecture in [20] that all sequences with almost perfect linear complexity profile can be obtained with the construction in [20].

This work is organized as follows: In the next section, we briefly recall how one is naturally lead from linearly recurrent sequences to expansion sequences. The third section is devoted to an analysis via continued fraction expansion. In the forth section, we give the theoretical background of our attack, and in the fifth section, we discuss computational aspects. The final section contains a discussion and some research proposals based on our attack.

Some definitions Let k be a finite field. By a *function field* over k we mean a finitely generated extension of k of transcendence degree 1. Let F/k be a function field with exact constant field k (which means that k is algebraically closed in F), and let $f \in F$ be non-constant (that is, $f \notin k$). Then we define the *degree* of f as $\deg(f) := [F : k(f)]$. Thus the degree of f is the degree of f as a function on the corresponding complete non-singular curve over k. In particular, the degree of a non-constant rational function $r = \frac{a}{b} \in k(t)$ with coprime polynomials $a, b \in k[t]$ is $\max\{\deg(a), \deg(b)\}$. Additionally, we define the *valuation degree* of r as $\operatorname{valdeg}(r) := \deg(a) - \deg(b)$. We therefore have $v_\infty(r) = -\operatorname{valdeg}(r)$. We caution the reader to not confuse the degree of a rational function with the valuation degree. The latter will only be used in Section 3 on continued fractions.

Acknowledgments I thank Andrei Shelest and Natalia Vasilevskaya for introducing me to the idea to generate random sequences by power series expansions and for discussions. I thank Enric Nart for a discussion on Hensel's Lemma.

2 From linearly recurrent sequences to power series expansions

Let us first recall some facts about linear complexity. These facts lead naturally to the consideration of continued fraction expansions and to the construction in [20].

Let still q be a prime power, and let $\mathbf{a} = (a_1, a_2, \ldots)$ be an infinite sequence over $\mathbb{F}_q$. Then we have the associated generating series

$$s := \sum_{i \in \mathbb{N}} a_i t^i \in \mathbb{F}_q[[t]] \ .$$

With $x := t^{-1} \in \mathbb{F}_q(t)$ we obtain that

$$s = \sum_{i \in \mathbb{N}} a_i x^{-i} \in \mathbb{F}_q[[x^{-1}]] \ .$$

Both these descriptions of the series s are of importance in the following.

The sequence $\mathbf{a}$ is generated by a linear feedback shift register with generating polynomial $g(x)$ if and only if $s \cdot g$ is a polynomial in x, which is then automatically of degree $< \deg(g)$. Therefore, $\mathbf{a}$ is linearly recurrent with recursion order d if and only if s is a rational function in x of degree d. Moreover, we have $L_n(\mathbf{a}) \leq \ell$ if and only if there exists a rational function $f \in \mathbb{F}_q(x)$ of degree at most ℓ with $s = f + O(x^{-(n+1)})$, where the O-notation is used as in infinitesimal calculus.

It is now natural to consider the continued fraction expansion of s to study the linear complexity profile of $\mathbf{a}$. This is done in [14] and in [16]. In the next section we recall some results obtained in this way and make some more observations. Here, we mention just one basic result which can be obtained in this way (see Proposition 2 in the next section):

As above, let $\mathbf{a}$ be an arbitrary infinite sequence over $\mathbb{F}_q$ and let $d \in \mathbb{N}$. Then $\mathbf{a}$ has d-almost perfect linear complexity profile if and only if $L_n(\mathbf{a}) \geq \frac{1}{2}(n + 1 - d)$ for all $n \in \mathbb{N}$.

Whereas the first description of the sequence s is important for the analysis via continued fraction expansion, it is the second description which leads naturally to the construction via function expansion into power series as described in [20] and also in [16]. We observe that s is a rational function of degree d in x if and only if it is a rational function of degree d in t. Therefore, s is linearly recurrent with recurrence order d if and only if s is a rational function of degree d in t.

Let now $F/\mathbb{F}_q$ be a function field with exact constant field $\mathbb{F}_q$, and $\mathfrak{P}$ be a place of degree 1 of F and t a uniformizing element at $\mathfrak{P}$. Then we have an associated homomorphism of function fields $F \hookrightarrow \mathbb{F}_q((t))$. The image of

some $f \in F$ is called the *expansion* of f with respect to t. Now for any f with $v_P(f) \geq 1$, we obtain a power series $\sum_{i\in\mathbb{N}} a_i t^i$, and we can consider the series $(a_i)_{i\in\mathbb{N}}$ defined by the coefficients.

For example, if $F = \mathbb{F}_q(t)$, that is, if t has degree 1, then we obtain a linearly recurrent sequence with recurrence order $\deg(f)$. By the above considerations, it is particularly natural here to consider expansions of functions $f \in \mathbb{F}_q(x)$ with respect to the "place at infinity" (with respect to x), $\mathfrak{p}_\infty$, and the local uniformizer $t := x^{-1}$.

It is proven in [20] that if t has degree 2 and f is in $F - \mathbb{F}_q(t)$ and has degree d then the sequence $(a_i)_{i\in\mathbb{N}}$ has d-almost perfect linear complexity profile. More generally, for arbitrary non-constant t we have

$$L_{\mathbf{a}}(n) \geq \frac{n+1-\deg(f)}{\deg(t)}$$

for all $n \in \mathbb{N}$.

Let us recall the easy proof: Let $\ell := L_{\mathbf{a}}(n)$. Then we have polynomials $r, u \in \mathbb{F}_q[t]$ with $\deg(r) \leq \ell, \deg(u) \leq \ell, u(0) = 0$ and

$$s \cdot r \equiv u \mod t^{n+1} .$$

This is equivalent to

$$f \cdot r \equiv u \mod \mathfrak{P}^{n+1} ,$$

i.e.

$$v_{\mathfrak{P}}(f \cdot r - u) \geq n+1 .$$

This implies that $\deg(f\cdot r-u) = \deg(f\cdot r-u)_0 \geq n+1$. On the other hand, $f\cdot r-u \in L((f)_\infty+\ell\cdot(t)_\infty)-\{0\}$ and thus $\deg(f\cdot r-u) \leq \deg(f)+\ell\cdot\deg(t)$. We conclude that $n+1 \leq \deg(f)+\ell\cdot\deg(t)$.

Definition We call a power series in $\mathbb{F}_q((t))$ obtained by expansion of any function as just described an *expansion series* (over $\mathbb{F}_q$). Furthermore, if $\mathbf{a}$ is a series which is the sequence of coefficients of an expansion series (which then lies in $t\cdot\mathbb{F}_q[[t]]$), then we call the sequence $\mathbf{a}$ an *expansion sequence* (over $\mathbb{F}_q$).

3 Analysis with continued fraction expansion

Again let q be a prime power. We consider continued fraction expansions of elements of $\mathbb{F}_q((x^{-1}))$. We expend the valuation degree function from $\mathbb{F}_q(x)$ to $\mathbb{F}_q((x^{-1}))$ by $\operatorname{valdeg}(\sum_{i\geq n} a_i x^{-i}) := -n$ if $a_n \neq 0$, and we set $v_\infty(f) := -\operatorname{valdeg}(f)$ and $|f| := q^{-v_\infty(f)} = q^{\operatorname{valdeg}(f)}$. Clearly, $|\cdot|$ is a non-archimedean absolute value, and $\mathbb{F}_q((x^{-1}))$ is the completion of $\mathbb{F}_q(x)$ with respect to this absolute value.

Let now $\mathbf{a}$ be an infinite sequence over $\mathbb{F}_q$. We consider the continued fraction expansion of $s = \sum_{i\in\mathbb{N}} a_i x^{-i}$. Let us first fix some standard definitions and recall some basic results. We mainly follow [16] here.

Let

$$s = [0; A_1, A_2, A_3, \ldots]$$

be the continued fraction expansion of s. Here by definition, the A_i are polynomials in $\mathbb{F}_q[x]$; they are called the *partial quotients* of s.

As usual, we define

$$p_{-1} := 1\,,\ p_0 := 0\,,\ p_i := A_i p_{i-1} + p_{i-2} \quad \text{for } i \in \mathbb{N}$$

and

$$q_{-1} := 0\,,\ q_0 := 1\,,\ q_i := A_i q_{i-1} + q_{i-2} \quad \text{for } i \in \mathbb{N}\,.$$

Then for each i, the polynomials p_i and q_i are coprime, and we have

$$\frac{p_i}{q_i} = [0; A_1, \ldots, A_i]$$

and

$$\deg(q_i) = \sum_{j\le i} \deg(A_i)\,.$$

We set $w_i := \deg(q_i)$.

In [14] the following proposition is proven (see also [16]).

Proposition 1 *Let $n \in \mathbb{N}$. Let now j be defined by the following inequalities*

$$w_{j-1} + w_j \le n < w_j + w_{j+1}\,. \tag{1}$$

Then $L_n(\mathbf{a}) = w_j$.

A consequence of this proposition is:

Proposition 2 *Let $d \in \mathbb{N}$. Then the following conditions are equivalent:*

a) The sequence $\mathbf{a}$ has d-almost regular complexity profile.

b) $L_n(\mathbf{a}) \le \frac{n+d}{2}$ for all $n \in \mathbb{N}$.

c) $L_n(\mathbf{a}) \ge \frac{n+1-d}{2}$ for all $n \in \mathbb{N}$.

d) $\deg(A_i) \le d$ for all i.

Later we will consider a variant of this proposition. For this reason, we now recall the proof given in [16].

Clearly, a) implies b) and c). We show that b) implies d) and that c) implies d).

Let $i \in \mathbb{N}$ with $w_i \geq 1$ and let $n := w_{i-1} + w_i$. Then $L_n(\mathbf{a}) = w_i$ and $n = 2w_i - \deg(A_i) = 2L_n(\mathbf{a}) - \deg(A_i)$. Therefore $\deg(A_i) = 2L_n(\mathbf{a}) - n$. If now b) is satisfied then $\deg(A_i) \leq d$.

Let now $i \in \mathbb{N}$ with $\deg(A_i) \geq 2$ and let $n := w_{i-1} + w_i - 1$. Then $L_n(\mathbf{a}) = w_{i-1}$ and $n = 2w_{i-1} + \deg(A_j) - 1 = 2L_n(\mathbf{a}) + \deg(A_j) - 1$. Therefore $\deg(A_i) = n + 1 - 2L_n(\mathbf{a})$. If now c) is satisfied then again $\deg(A_i) \leq d$.

We now show that d) implies b) and c) and therefore also a).

Let $n \in \mathbb{N}$. Then inequalities (1) are equivalent to

$$2w_j - \deg(A_j) \leq n < 2w_j + \deg(A_j) \ ,$$

which is equivalent to

$$n - \deg(A_j) + 1 \leq 2w_j \leq n + \deg(A_j) \ .$$

The claim follows immediately. □

A remark on expansion sequences Let again $d \in \mathbb{N}$. We see from the previous proposition that there are uncountably many sequences over $\mathbb{F}_q$ with d-almost perfect complexity profile. On the other hand, there are only countably many expansion sequences, even if one does not require that the degree of the function is 2. In contrast to this, in the conclusion of [20] it is conjectured that all sequences with almost perfect complexity profile are expansion sequences of functions of degree 2. We see that this conjecture fails in a dramatic way.

It is now natural to study sequences obtained with the construction in [20] via the theory of continued fraction expansion. The following proposition is classical.

Proposition 3 *The following statements are equivalent:*

a) $\mathbf{a}$ *is an expansion sequence of a function* f *in a function field* F *with exact constant field* $\mathbb{F}_q$ *with respect to a place of degree 1 and a uniformizing parameter* t *of degree 2, where* $f \notin \mathbb{F}_q(t)$.

b) There exists a quadratic field extension $F|\mathbb{F}_q(x)$ *for which the place* $\mathfrak{p}_\infty$ *of* $\mathbb{F}_q(x)$ *is unramified and split into two places* $\mathfrak{P}_1, \mathfrak{P}_2$ *of* F *and there exists a function* $f \in F - \mathbb{F}_q(x)$ *such that* s *is the expansion sequence of* f *at* $\mathfrak{P}_1$ *with respect to the uniformizing parameter* x^{-1}.

c) s *is a root of an irreducible polynomial in* $\mathbb{F}_q(x)[y]$ *of degree 2 in* y.

d) The continued fraction expansion of s *is periodic.*

In analogy to quadratic number fields, for a field k, a quadratic extension $F|k(x)$ in which the place $\mathfrak{p}_\infty$ is unramified and split is often called a *real quadratic function field*. Note here that this might be seen as an abuse of terminology as a "real quadratic function field" is not a function field but an extension of the field $\mathbb{F}_q(x)$. Just as the theory of continued fractions for real quadratic number fields, the corresponding theory for real quadratic function fields over finite fields is well developed. A good overview over many aspects for odd characteristic is [19], the case of even characteristic is discussed in [21].

The theory of continued fraction can be used to obtain bounds on the degrees of the partial fractions A_i. We now discuss these results and relate them to Proposition 2.

Let now $ay^2 + by + c \in \mathbb{F}_q[x, y]$ with $a, b, c \in \mathbb{F}_q[x]$ and $a \neq 0$ be an irreducible polynomial. Let F be the extension of $\mathbb{F}_q(x)$ defined by the polynomial, and let f the residue class of y. We assume that $\mathfrak{p}_\infty$ is unramified and split in F; let $\mathfrak{P}_1$ be one of these places. We consider the expansion s of f at $\mathfrak{P}_1$ and the corresponding continued fraction expansion. We use the notations from above.

We note first that

$$\deg(f) = \max\{\deg(a), \deg(b), \deg(c)\}$$

and therefore

$$\deg(A_i) \leq \max\{\deg(a), \deg(b), \deg(c)\} \tag{2}$$

for all $i \in \mathbb{N}$ by the considerations of the previous section and Proposition 2. We now give potentially better bounds for i large enough.

The function $g := af$ is a root of the monic polynomial $y^2 + by + ac$, so af is integral over $\mathbb{F}_q[x]$. Moreover, clearly, a divides ac which is the norm of g over $\mathbb{F}_q(x)$. We therefore have

$$f = \frac{g}{a}$$

with g integral, $a \in \mathbb{F}_q[x]$ and $a|N(g)$. We are now in the situation which is considered in continued fraction expansions.

We make a case distinction according to whether the characteristic is odd or even.

We first consider the "classical" case that the characteristic is *odd*. We have the discriminant $\Delta = b^2 - 4ac$. From Proposition 3.2 (c) of [19] we learn that $|A_i| \leq |\sqrt{\Delta}|$, that is, $|A_i|^2 \leq |\Delta|$, for i large enough. In other words:

$$\deg(A_i) \leq \frac{\deg(\Delta)}{2}$$

for i large enough. This inequality is always at least as strong as inequality (2). By the proof of Proposition 2, we have

$$n - \frac{\deg(\Delta)}{2} + 1 \leq 2L_n(\mathbf{a}) \leq n + \frac{\deg(\Delta)}{2}$$

for n large enough.

Let now the characteristic be even. Then by Section 3 of [21] we have $|A_i| \leq |b|$, that is,

$$\deg(A_i) \leq \deg(b)$$

for i large enough. (In the introduction to [21] there are some assumptions on the minimal polynomial of g (in our notation), but these assumptions are not relevant for Section 3 of [21].) We thus have

$$n - \deg(b) + 1 \leq 2L_n(\mathbf{a}) \leq n + \deg(b)$$

for n large enough.

4 Defining polynomials

Let an expansion sequence $\mathbf{a}$ over $\mathbb{F}_q$ defined by some function field F with exact constant field $\mathbb{F}_q$ a place $\mathfrak{P}$ of degree 1 and a uniformizing parameter t be given. Note here that we make no assumption on the degree of t.

Convention Let us assume that $\mathbb{F}_q(t, f)$ is a proper subfield of F. Let now $\mathfrak{P}'$ be the restriction of $\mathfrak{P}$ to $\mathbb{F}_q(t, f)$. Then $\mathfrak{P}'$ is also a place of degree 1, and the series of f defined by $\mathfrak{P}$ and the local parameter t is identical to the one defined by $\mathfrak{P}'$ and the local parameter t. So, in our study of expansion sequences, we can restrict our attention to sequence arising as above with $F = \mathbb{F}_q(t, f)$, and we do so in the following.

The elements t and f are algebraically dependent over $\mathbb{F}_q$. So there exists a non-trivial polynomial $h = h(t, y) \in \mathbb{F}_q[t, y]$ with

$$h(t, f) = 0\,. \tag{3}$$

Equation (3) is equivalent to

$$h(t, \sum_{i \in \mathbb{N}} a_i t^i) = 0\,, \tag{4}$$

so we have a non-trivial polynomial h which satisfies the latter condition. Let now I be the ideal of polynomials $h \in \mathbb{F}_q[t, y]$ with $h(t, f) = 0$.

The ideal I is the kernel of the homomorphism

$$\mathbb{F}_q[t, y] \longrightarrow F\,,\ t \mapsto t\,,\ y \mapsto f \tag{5}$$

and also of the homomorphism

$$\mathbb{F}_q[t,y] \longrightarrow \mathbb{F}_q[[t]]\ ,\ t \mapsto t\ ,\ y \mapsto \sum_{i\in\mathbb{N}} a_i t^i\ . \tag{6}$$

Note that the latter fact implies that I is canonically attached to the sequence $\mathbf{a}$.

The ideal I is a prime ideal and $V(I)$ is a (possibly singular) plane affine curve with function field F. As $V(I)$ is one-dimensional, by Krull's Hauptidealsatz I is generated by a single irreducible polynomial. As usual, we call such a polynomial a *defining polynomial* of the curve $V(I)$. Such a polynomial is unique up to a constant.

Let us consider the situation from the point of view of projective geometry: Let for this $\mathcal{C}$ be the complete non-singular curve over $\mathbb{F}_q$ corresponding to F, where we fix an isomorphism of function fields $\mathbb{F}_q(\mathcal{C}) \simeq F$. The place $\mathfrak{P}$ corresponds to an $\mathbb{F}_q$-rational point of $\mathcal{C}$ which we denote by P. Let now $h_0(t,y)$ be a defining polynomial of $V(I)$, and let $H_0(T,Y,Z) \in \mathbb{F}_q[T,Y,Z]$ be the homogenization of h_0. Let

$$D_0 := \sup\{-\operatorname{div}(t), -\operatorname{div}(f), 1\} = \sup\{(t)_\infty, (f)_\infty\}\ , \tag{7}$$

where for some function $g \in F^*$, $(g)_\infty$ is the pole divisor of g. Clearly, $\deg(D_0) \le \deg(t) + \deg(f)$. By definition, $t, f, 1$ generate $\mathcal{O}(D_0)$. Thus $\mathcal{O}(D_0)$ and the global sections $t, f, 1$ of $\mathcal{O}(D_0)$ define a morphism $\mathcal{C} \longrightarrow \mathbb{P}^2_{\mathbb{F}_q}$ which is given on an open subset U of $\mathcal{C}$ by $Q \mapsto (t(Q) : f(Q) : 1)$ for $Q \in U(\overline{\mathbb{F}}_q)$. (Every rational map from $\mathcal{C}$ to some projective space $\mathbb{P}^n_{\mathbb{F}_q}$ over $\mathbb{F}_q$ can be extended to a morphism from $\mathcal{C}$ to $\mathbb{P}^n_{\mathbb{F}_q}$. Here we have the stronger condition that $t, f, 1$ as global sections of $\mathcal{O}(D_0)$ directly define a morphism. See Section II.7 of [8] for background information.)

As by assumption $F = \mathbb{F}_q(t,f)$, the morphism is birational onto its image. It follows that the image $V(H_0)$ has degree $\deg(D_0)$. We have therefore

$$\deg(h_0) = \deg(H_0) = \deg(D_0) \le \deg(t) + \deg(f)\ . \tag{8}$$

We now turn the situation around and just assume that we are given some sequence $\mathbf{a}$ over $\mathbb{F}_q$ for which a non-trivial polynomial $h \in \mathbb{F}_q[t,y]$ with $h(t, \sum_{i\in\mathbb{N}} a_i t^i) = 0$ exists. Then the polynomials h with this property define again a non-trivial proper ideal of $\mathbb{F}_q[t,y]$; let $I_{\mathbf{a}}$ be this ideal. Just as above, let h_0 be a polynomial of $I_{\mathbf{a}}$ of minimal degree. Clearly, h_0 is irreducible.

Let $F := \mathbb{F}_q(t)[y]/(h_0)$ and f the residue class of y. Then we have the embedding $F \longrightarrow \mathbb{F}_q((t))$ given by $t \mapsto t$ and $f \mapsto \sum_{i\in\mathbb{N}} a_i t^i$ over $\mathbb{F}_q$. This embedding defines in a unique way a valuation v on F with $v(t) = 1$ and

thus a place $\mathfrak{P}$ of F of degree 1. Moreover, the expansion of f at $\mathfrak{P}$ with respect to t is of course the power series $\sum_{i\in\mathbb{N}} a_i t^i$.

The ideal $I_{\mathbf{a}}$ is now by definition the kernel of the homomorphism in (5) and thus also the kernel of the homomorphism in (6). By the previous considerations we conclude that $I_{\mathbf{a}}$ is generated by h_0.

We have proven:

Proposition 4 *The expansion sequences are exactly the sequences* $\mathbf{a}$ *for which a non-trivial polynomial* $h \in \mathbb{F}_q[t,y]$ *with* $h(t, \sum_{i\in\mathbb{N}} a_i t^i) = 0$ *exists. Furthermore, if such a polynomial exists, the ideal of such polynomials is generated by a single irreducible polynomial.*

This leads to the following definition.

Definition Let $\mathbf{a}$ be an expansion sequence. We call the ideal $I_{\mathbf{a}}$ the *defining ideal* of $\mathbf{a}$, any non-trivial element of $I_{\mathbf{a}}$ a *defining polynomial* of $\mathbf{a}$ and a generating element h_0 of $I_{\mathbf{a}}$ a *minimal defining polynomial* of $\mathbf{a}$. Finally, we call the degree of a minimal defining polynomial of $\mathbf{a}$ the *degree* of $\mathbf{a}$ (as an expansion sequence).

Note that a necessary condition that a polynomial $h_0 \in \mathbb{F}_q[t,y]$ is a defining polynomial of any expansion sequence is that its constant term is trivial.

Let now h_0 be an irreducible polynomial in $\mathbb{F}_q[t,y]$ with trivial constant term. As above, let $F = \mathbb{F}_q(t)[y]/(h_0)$, and let f be the residue class of y.

Now the expansion sequences with minimal defining polynomial h_0 correspond bijectively to the places $\mathfrak{P}$ of F of degree 1 with $t, f \equiv 0 \bmod \mathfrak{P}$. From a geometric point of view, the situation is as follows: Let H_0 be the homogenization of h_0 and $(\mathcal{C}, \varphi)$ a non-singular curve $\mathcal{C}$ over $\mathbb{F}_q$ together with a birational morphism $\varphi : \mathcal{C} \longrightarrow V(H_0)$; this datum is unique up to unique isomorphism. Now the places $\mathfrak{P}$ of F of degree 1 with $t, f \equiv 0 \bmod \mathfrak{P}$ correspond in a unique way to points $P \in \mathcal{C}(\mathbb{F}_q)$ with $\varphi(P) = [0:0:1]$.

By this geometric description, two facts are immediate:

First, there are at most $\deg(h_0)$ expansion sequences with minimal defining polynomial h_0.

Second, if $V(h_0)$ is non-singular at $(0,0)$ then there is exactly one such sequence. Let now $h_0 = \sum_{i,j} c_{i,j} t^i y^j$. Then h_0 is singular at $(0,0)$ if and only if $c_{1,0} = c_{0,1} = 0$.

For the general case, we have the following proposition.

Lemma 5 *An expansion sequence of degree* d_0 *is uniquely determined by its defining polynomial and its initial sequence of length* d_0^2.

Proof. We consider h_0 as a polynomial in y. As t is a local parameter of $\mathfrak{P}$, the extension $F|\mathbb{F}_q(t)$ is separable, and therefore h_0 is a separable polynomial. Let $h_0 = \sum_{i=0}^{d} c_i(t) \cdot y^i$ with $c_i(t) \in \mathbb{F}_q[t]$ and $c_d(t) \neq 0$, let Δ be the discriminant of h_0, and let R be the resultant of h_0 and $h_0' = \frac{\partial h_0}{\partial y}$. We have $R = \pm c_d \, \Delta$, R is a non-trivial polynomial in $\mathbb{F}_q[t]$ of degree at most $d_0(2d_0 - 1)$, and the discriminant Δ is a non-trivial polynomial in $\mathbb{F}_q[t]$ too.

Assume now that we have two distinct roots $\rho, \tilde{\rho}$ of h_0 as a polynomial over $\mathbb{F}_q[[t]]$ with $\rho \equiv \tilde{\rho} \bmod t^{n+1}$ for some n. Then $\rho - \tilde{\rho} \equiv 0 \bmod t^{n+1}$. Now $(\rho - \tilde{\rho})^2$ is a divisor of Δ. Therefore $\Delta \equiv 0 \bmod t^{2n+2}$. We obtain that $2n + 2 \leq d_0(2d_0 - 1)$, and therefore $n < d_0^2$. □

So far, we have fixed the minimal defining polynomial. The minimal defining polynomial of an expansion sequence of degree d_0 is however also uniquely determined by its initial sequence of degree d_0^2. The key statement is the following lemma. The proof of the lemma is an easy adaption of the proof of Lemma 2.2 in [1].

Lemma 6 *Let* $\mathbf{a}$ *be an expansion sequence over* $\mathbb{F}_q$ *of degree* d_0*. Let* $h \in \mathbb{F}_q[t, y]$ *be a non-trivial polynomial such that with* $d := \deg(h)$ *and* $n := dd_0$ *we have*

$$h(t, \sum_{i=1}^{n} a_i t^i) \equiv 0 \bmod t^{n+1} \, . \tag{9}$$

Then h *is a defining polynomial for* $\mathbf{a}$*.*

Proof. Let $\mathbf{a}$ be an expansion sequence defined by $F, f, t, \mathfrak{P}$, where these objects are as above. Let D_0 be as in (7). Then $h(t, f) \in L(d \cdot D_0)$. If now $h(t, f)$ is non-trivial then $\deg(h(t, f)) = \deg(h(t, f)_\infty) \leq \deg(h) \cdot \deg(D_0) = dd_0$.

Let us assume that congruence (9) holds for some $n \in \mathbb{N}$. Then $h(t, f) \equiv 0 \mod \mathfrak{P}^{n+1}$. Under the assumption that $h(t, f)$ is non-trivial, we now have $\deg(h(t, f)) \geq v_P(h(t, f)) \geq n + 1$.

If $n \geq dd_0$ this is a contradiction. Therefore, we then have $h(t, f) = 0$ and thus $h(t, \sum_{i \in \mathbb{N}} a_i t^i) = 0$. □

Together these two lemmata give the following result.

Proposition 7 *An expansion sequence of degree* d_0 *is uniquely determined by its initial sequence of length* d_0^2*.*

5 Computations

We now discuss computational aspects.

Our main goal is here as follows: Given an initial sequence of length at least d_0^2 of an expansion sequence $\mathbf{a}$ of degree d_0, we want to compute large initial sequences of $\mathbf{a}$. For this, we proceed as follows: First, we compute a minimal defining polynomial. Then from the minimal defining polynomial and the initial sequence we compute the further coefficients of the expansion sequence.

At the end of the section we also discuss why the results also apply if a finite intermediate sequence is given. This case is particularly relevant for cryptanalytic applications.

Recovering the minimal defining polynomial

Lemma 6 immediately gives rise to an algorithm to determine a minimal defining polynomial from a suitable initial sequence: Let some finite initial sequence of length at least d_0^2 of an expansion sequence of degree d_0 be given. The degree d_0 itself need not be known a priori.

For $d = 1, 2, \ldots$, we make an ansatz for h_0 as a polynomial of degree d with unknown coefficients. Then (9) with $n = d^2$ gives d^2 homogeneous linear equations on the coefficients of h. We solve this system of equations by linear algebra. If we have found a non-trivial solution, we compute the corresponding polynomial. This is then a minimal defining polynomial.

Note that the number of monomials in $\mathbb{F}_q[t, y]$ of degree at most d is $\binom{d+2}{2}$. Thus a particular system to be solved has size $d^2 \times \binom{d+2}{2}$. We therefore obtain the following proposition.

Proposition 8 *A minimal defining polynomial of an expansion sequence of degree d_0 can be computed in polynomial time in $d_0 \cdot \log(q)$ from an initial sequence of length d_0^2 or more.*

We remark that in practice, one might try to apply this idea with n at least $\binom{d+2}{2}$ but smaller than d^2.

Approaches to compute initial sequences

Let us now assume that we are given an initial sequence of length $n \geq d_0^2$ and a minimal defining polynomial h_0 of an expansion sequence $\mathbf{a}$ of degree d_0. We wish to compute efficiently further coefficients of the sequence $\mathbf{a}$. Three approaches to this problem come to mind:

1. A direct approach. For some $m > n$, we use the congruence $h_0(t, \sum_{i=1}^{m} a_i t^i) \equiv 0 \bmod t^{m+1}$ to obtain a system of equations for the unknowns $a_{n+1}, \ldots, a_m$. We then solve this system.

2. Expansions of functions with a function field theoretic approach. Let $F := \mathbb{F}_q[t, y]/(h_0)$, and let f be the residue class of y. Now every place $\mathfrak{P}$ of F of degree 1 with $t, y \equiv 0 \bmod \mathfrak{P}$ determines uniquely an expansion sequence of F, and one of these is $\sum_{i\in\mathbb{N}} a_i t^i$. By (5), the series $\mathbf{a}$ determines a unique place. We therefore have the following approach: First we determine all places $\mathfrak{P}$ with $t, y \equiv 0 \bmod \mathfrak{P}$. For each such place $\mathfrak{P}$, we compute the expansion of f at $\mathfrak{P}$ to the power d_0^2. From this, we determine which place is the correct one. Then we compute further coefficients as desired.

3. Hensel's Lemma. We use a non-archimedean variant of Newton's iteration or – with other words – we use some effective version of Hensel's Lemma.

The use of Hensel's Lemma (as formulated in [11]) to compute expansion sequences was already suggested in [10] for the case that t has degree 2. It is extremely efficient. However, in order that it can be applied, the initial sequence has to satisfy a condition (see condition (13) below). This condition are missing in [10].

Because Hensel's Lemma cannot always be applied and also because of independent interest, we now first discuss the function field theoretic approach. Then we come to Hensel's Lemma and give a criterion under which (11) is satisfied. The direct approach is not discussed in the following.

Let us for the following fix a definition:

Definition Let $\sum_{i\in\mathbb{N}_0} a_i t^i \in \mathbb{F}_q[[t]]$ and $n \in \mathbb{N}_0$. Then we call the polynomial $\sum_{i=0}^{n} a_i t^i$ the *initial series* of $\sum_{i\in\mathbb{N}_0} a_i t^i$ of *length* n.

Expansions of functions

We consider the general problem to compute an initial series of an expansion series of a function in a function field with respect to a place of degree 1 and a local parameter. For concreteness we only consider function fields over finite fields.

A first problem is how to represent the objects for computational purposes, most importantly the place of degree 1.

There are several approaches here, but one approach has proven itself to be particularly successful. The general idea of this approach is to adapt ideas which are successfully used for number fields. This approach has in particular been popularized by F. Hess in his work [9]. It is also implemented in the computer algebra system MAGMA ([2]). We describe this approach briefly. Besides [9], more information on this approach can be found in [6], and even more information, including proofs of all the following claims can be found in Chapter 2 of [5].

Let $F/\mathbb{F}_q$ be a function field, and let $\mathcal{C}$ be a non-singular proper curve over $\mathbb{F}_q$ with a fixed isomorphism $\mathbb{F}_q(\mathcal{C}) \simeq F$. (As we do not assume that $\mathbb{F}_q$ is the exact constant field of F here, the curve $\mathcal{C}$ need not be geometrically irreducible.) The field F itself shall be given by an irreducible polynomial $h_0 \in \mathbb{F}_q[t, y]$ which is separable in y. ($\mathcal{C}$ is then birational to $V(h_0)$.) We consider the separable field extension $F|\mathbb{F}_q(t)$. We have the "finite" order, which is the closure of $\mathbb{F}_q[t]$ in F and the "infinite" order, which is the closure of $\mathbb{F}_q[\frac{1}{t}]_{(\frac{1}{t})}$ in F. Bases of the orders over the base rings $\mathbb{F}_q[t]$ respectively $\mathbb{F}_q[\frac{1}{t}]_{(\frac{1}{t})}$ can be computed in polynomial time in $\deg(h_0) \cdot \log(q)$. Now a divisor on $\mathcal{C}$ is represented by a pair of two fractional ideals – one fractional ideal for each order. The fractional ideals themselves are also represented by bases over the base rings $\mathbb{F}_q[t]$ respectively $\mathbb{F}_q[\frac{1}{t}]_{(\frac{1}{t})}$. Basic arithmetic can now be performed just as for ideals in number fields and as described for example in [3]. Furthermore, as shown in [9], there is an easy algorithm to compute, for a given divisor D of $\mathcal{C}$, the Riemann-Roch space $L(D)$. With appropriate representations of the objects involved (which we have not described in all details), divisor arithmetic and the computation of the Riemann-Roch spaces can be performed in polynomial time in $\deg(h_0)$, $\log(q)$ and the degrees of the divisors involved.

We now come to our application. Let an irreducible polynomial $h_0 \in \mathbb{F}_q[t, y]$ which is separable in y be given. Let $F := \mathbb{F}_q[t, y]/(h_0)$. We are interested in the places $\mathfrak{P}$ of F of degree 1 with $t, y \equiv 0 \bmod \mathfrak{P}$. Let $D := \min\{(t)_0, (y)_0\}$. Then the places we search for are exactly the places of degree 1 in the support of D. Now, D can be computed in polynomial time in $\deg(h_0) \cdot \log(q)$. The factorization of D as a formal sum of prime divisors (the so-called "free representation" of D) can be computed in an expected time which is polynomially bounded in $\deg(h_0) \cdot \log(q)$. (Here we have to factor polynomials, so this part of the algorithm is randomized.) From this factorization, the desired places can be read off.

At this point, we are left with the task to compute an initial part of an expansion series at a place of degree 1 with respect to the local parameter t. We consider this from a more general point of view.

We consider the general situation described above. So let again F be a function field over $\mathbb{F}_q$, which is now given by some irreducible polynomial $h_0 \in \mathbb{F}_q[t, y]$ which is separable in y. Let $\mathcal{C}$ be a complete non-singular curve with a fixed isomorphism $F \simeq \mathbb{F}_q(\mathcal{C})$. Furthermore, let $\mathfrak{P}$ be a place of degree 1 of F with local parameter t. Let $P \in \mathcal{C}(\mathbb{F}_q)$ be the point corresponding to the place $\mathfrak{P}$.

Let now f_i for $i \in \mathbb{N}_0$ be inductively defined as follows:

$$f_0 := f \quad , \quad f_{i+1} := \frac{f_i - f_i(P)}{t} \tag{10}$$

Then the expansion sequence of f at $\mathfrak{P}$ with respect to t is $\sum_{i=1}^{\infty} a_i t^i$ with $a_i = f_i(P)$.

So we only have to consider one algorithmic problem: How can one efficiently evaluate a function at an $\mathbb{F}_q$-rational point? That is, given F, f and P, how can one compute $f(P)$?

Let us first mention that this problem is not completely trivial for various reasons. One reason is that the point is not given by t and y-coordinates but as an ideal in an order of F. But let us now assume that we know the coordinates $t(P), y(P)$. If then f is given by a polynomial in $\mathbb{F}_q[t, y]$, it is trivial to evaluate f at P. If however f is given as a fraction of polynomials or a sum of fractions of polynomials, it is not a priori clear now to perform the evaluation in complete generality.

There is however an easy solution via Riemann-Roch spaces: Note first that $f(P)$ is the unique element $a \in \mathbb{F}_q$ such that $f - a$ vanishes at $\mathfrak{P}$. All functions $f - a$ lie in $L((f)_\infty)$, and they lie in $L((f)_\infty - P)$ if and only if $a = f(P)$.

So we first compute a basis $b_1, \ldots, b_\ell$ of the space $L((f)_\infty - P)$. Then $1, b_1, \ldots, b_\ell$ is a basis of $L((f)_\infty)$. We determine $a, a_1, \ldots, a_\ell \in \mathbb{F}_q$ with $f = a + a_1 b_1 + \cdots + a_\ell b_\ell$ with a linear algebra computation. Then $f - a \in L((f)_\infty - P)$ and therefore $f(P) = a$. This computation can be performed in polynomial time in $\deg(f) \cdot \deg(h_0) \cdot \log(q)$.

Now f_k lies in $L((f)_\infty + k \cdot (t)_\infty)$ and therefore $\deg(f_k) \leq \deg(f) + k \cdot \deg(t) \leq (k+1) \cdot \deg(h_0)$. It follows that the computation of the initial series $\sum_{i=1}^{n} a_i t^i$ can be performed in polynomial time in $n \cdot \deg(h_0) \cdot \log(q)$.

Let us now return to our initial problem, the computation of an sequence of a particular length n from an initial sequence of length d_0^2, where d_0 is the degree of the expansion sequence.

We have to consider expansions with respect to t at all $\mathbb{F}_q$-rational points P of $\mathcal{C}$ with $(t(P), y(P)) = (0, 0)$. There are $\leq d_0$ such points. Therefore, the computation can be performed in polynomial time in $n \cdot d_0 \cdot \log(q)$.

Hensel's Lemma

The computation via Hensel's Lemma is particularly efficient – provided that it is possible.

The following lemma and proposition are crucial.

Lemma 9 *Let R be a complete discrete valuation ring with normalized valuation v, and let t be a local parameter. Let $g \in R[y]$, and let $r \in R$ be such that $v(g(r)) > 2v(g'(r))$. Let*

$$\tilde{r} := r - \frac{g(r)}{g'(r)} \ .$$

Then

$$v(g'(\tilde{r})) = v(g'(r)) \quad , \quad v(g(\tilde{r})) - 2v(g'(\tilde{r})) \geq 2(v(g(r)) - 2v(g'(r))) \; .$$

This lemma is proven in the course of the proof of Proposition 2 in Section II.2 of [11]. The next proposition follows easily.

Proposition 10 *Let R be a complete discrete valuation ring with normalized valuation v, and let t be a local parameter. Let $g \in R[y]$, and let $r_0 \in R$ be such that*

$$v(g(r_0)) > 2v(g'(r_0)) \; . \tag{11}$$

Let $b := v(g(r_0)) - 2v(g'(r_0))$. Now let $(r_k)_{k\in\mathbb{N}}$ be a sequence in R such that for $k \in \mathbb{N}_0$

$$r_{k+1} \equiv r_k - \frac{g(r_k)}{g'(r_k)} \mod t^{2^{k+1}b+2v(g'(r_0))} \; . \tag{12}$$

Then

$$v(g'(r_k)) = v(g'(r_0)) \quad , \quad v(g(r_k)) - 2v(g'(r_k)) \geq 2^k \cdot b \; ,$$

$$r_{k+1} \equiv r_k \mod t^{b\cdot 2^k + v(g'(r_0))} \quad , \quad g(r_k) \equiv 0 \mod t^{b\cdot 2^k + 2v(g'(r_0))}$$

for all $k \in \mathbb{N}_0$. In particular, $(r_k)_{k\in\mathbb{N}_0}$ converges to a root ρ of g with $\rho \equiv r_0 \mod t^{v(g(r_0))-v(g'(r_0))}$ and, more generally, $\rho \equiv r_k \mod t^{b\cdot 2^k+v(g'(r_0))}$. Moreover, ρ is the unique root of g with $\rho \equiv r_0 \mod t^{v(g'(r_0))+1}$ (and thus in particular with $\rho \equiv r_0 \mod t^{v(g(r_0))-v(g'(r_0))}$).

The proposition is one of the various statements which might be called "Hensel's Lemma". It is closely related to Proposition 2 in Section II.2 of [11]. However, we only require that the congruence (12) is satisfied. In contrast in Section II.2 of [11] an equality in R is demanded.

For the *proof*, note first that the congruence in the conclusion (except the uniqueness) follow immediately from the first two statements and (12).

The statements for arbitrary k follow immediately by induction from the lemma. The induction base $k = 0$ is trivial. So let us assume that the statements hold for a particular natural number k. Let $\tilde{r}_{k+1} := r_k - \frac{g(r_k)}{g'(r_k)}$. Then by the lemma applied with r_k and $\tilde{r}_{k+1}$ we obtain that $v(g'(\tilde{r}_{k+1})) = v(g'(r_k)) = v(g'(r_0))$ and $v(g(\tilde{r}_{k+1})) - 2v(g'(\tilde{r}_{k+1})) \geq 2(v(g(r_k)) - 2v(g'(r_k))) \geq 2^{k+1}b$, that is, $v(g(\tilde{r}_{k+1})) \geq 2^{k+1}b + 2v(g'(r_0))$. As $r_{k+1} \equiv \tilde{r}_{k+1} \mod t^{2^{k+1}b+2v(g'(r_0))}$ it follows that $v(g(r_{k+1})) \geq 2^{k+1}b + 2v(g(r_0))$ and $v(g'(r_k)) = v(g'(\tilde{r}_k)) = v(g'(r_0))$.

The uniqueness is essentially stated in Proposition 4.1.37 of [4]. For the convenience of the reader, we recall the easy proof here:

Let us assume that there are two distinct roots $\rho_1, \rho_2 \in R$ of g with $\rho_1 \equiv \rho_2 \equiv r_0 \mod t^{v(g'(r_0))+1}$. By Gauß' Lemma we have $g = (y-\rho_1)(y-\rho_2)h$

with some $h \in R[y]$. This gives $g'(\rho_1) = (\rho_1 - \rho_2)h(\rho_1)$ and therefore $v(g'(r_0)) = v(g'(\rho_1)) \geq v(\rho_1 - \rho_2)$, that is, $\rho_1 \not\equiv \rho_2 \mod t^{v(g'(r_0))+1}$, a contradiction. □

Just as previously, let $h_0 \in \mathbb{F}_q[t, y]$ be an irreducible polynomial, let $F := \mathbb{F}_q[t, y]/(h_0)$, and let f be the residue class of y. We which to apply the above "Hensel's Lemma" with $R = \mathbb{F}_q[[t]]$, v the corresponding normalized valuation, $g = h_0$ and $r_0 = \sum_{i=1}^{n} a_i t^i$ for some initial sequence $(a_1, \ldots, a_n)$.

We now address the task to find a suitable condition under which

$$v(h_0(r_0)) > 2v(\frac{\partial h_0}{\partial y}(r_0)) \tag{13}$$

holds.

Note first that the condition is satisfied if $\frac{\partial h_0}{\partial y}(r_0) \neq 0$. We now consider the general case.

Proposition 11 *Let* **a** *be an expansion sequence of degree d_0 with minimal defining polynomial $h_0 = \sum_{i=0}^{d} c_i(t)y^i$, and let $n \geq 2d_0^3$. Then $r_0 := \sum_{i=1}^{n} a_i t^i$ fulfills the condition $v(h_0(r_0)) > 2v(\frac{\partial h_0}{\partial y}(r_0))$.*

Proof. Let Z be the splitting field of $h_0 \in \mathbb{F}_q(t)[y]$ over $\mathbb{F}_q(t)$. We fix an embedding $F \hookrightarrow Z$, and we prolong the valuation v to Z. We denote the resulting valuation again by v.

Let $h_0 = c \cdot \prod_{j=1}^{m}(y - f_j) \in Z[y]$, where $c \in \mathbb{F}_q[t]$ and $f_1 = f$. By Gauß' Lemma, $v(f_j) \geq 0$ for all j. Furthermore $v(h_0(r_0)) \geq v(c) + v(r_0 - f) \geq v(c) + n + 1 \geq n + 1$. With an easy adaption of the proof of Lemma 5 we have for $j > 1$ $v(f - f_j) \leq d_0^2$ and thus $v(r_0 - f_j) \leq d_0^2$ too.

We have $\frac{\partial h_0}{\partial y}(r_0) = c \cdot \sum_{\ell=1}^{m} \prod_{j \neq \ell}(y - f_j)$. Now for $\ell > 1$, $v(\prod_{j \neq \ell}(r_0 - f_j)) \geq v(r_0 - f) \geq n + 1$. Furthermore $v(\prod_{j>1}(r_0 - f_j)) \leq (m - 1) \cdot d_0^2 \leq d_0^3 - d_0^2$. Therefore $v(\frac{\partial h_0}{\partial y}(r_0)) \leq v(c) + d_0^3 - d_0^2 \leq d_0^3$ and $2\frac{\partial h_0}{\partial y}(r_0) \leq 2d_0^3 < n + 1 \leq v(h_0(r_0))$. □

The computation is in principle straightforward: Let an irreducible polynomial h_0 such that (13) holds and a finite sequence $(a_1, \ldots, a_n)$ be given. Let us assume that (13) holds with $r_0 := \sum_{i=1}^{n} a_i t^i$.

We apply Hensel's Lemma with $g := h_0$. As above, let $b := v(g(r_0)) - 2v(g'(r_0))$. In the k-th iteration, we compute the unique polynomial of minimal degree r_{k+1} for which (12) holds. The computation takes place inside the residue class ring $\mathbb{F}_q[t]/(t^{2^{k+1}b+2v(g'(r_0))})$.

Note that this computation is very efficient because we essentially double the length of the computed initial sequence at every iteration. One can now combine this method with fast arithmetic. Like this, one obtains:

Proposition 12 *There exists a Turing machine with the following specification: Upon input of an irreducible polynomial h_0 such that (13) holds and a finite initial sequence $(a_1, \ldots, a_n)$ for which (13) holds the machine never terminates and it outputs the coefficients of an expansion sequence with minimal defining polynomial h_0 and initial sequence $(a_1, \ldots, a_n)$. Moreover, the running time until the m^{th} coefficient is output is in $\tilde{O}(m \cdot \deg(h_0))$.*

Storage requirements

With the two methods discussed above and also with the "direct approach" mentioned above, the storage requirements are enormous: In order to compute a_{n+1}, the complete initial sequence $(a_1, \ldots, a_n)$ has to be stored. It is an interesting question if there is any method to compute entries of $\mathbf{a}$ which uses less storage.

For the time being, the storage requirements put serious constraints on the use of expansion sequences for stream ciphers – independently of our attack.

Finite intermediate sequences

We now consider a variant of the above. In cryptanalytic applications, it is unlikely that one has access to an initial sequence. It is more realistic to assume that one has access to a finite subsequence $(a_{k+1}, \ldots, a_{k+m})$ of an expansion sequence $\mathbf{a}$.

Let $\mathbf{a}$ be given by $F, \mathfrak{P}, t, f$ as above, and furthermore let f_k be defined as in (10). Then the sequence $(a_{k+i})_{i \in \mathbb{N}}$ is the expansion sequence of $t \cdot f_{k+1}$ with respect to t at $\mathfrak{P}$. So all the above considerations hold when applied to this sequence.

Let now d_k be the degree of this expansion sequence. (This generalizes the definition of d_0 given above.) As $t \cdot f_{k+1} \in L((f)_\infty + k \cdot (t)_\infty)$, we have

$$d_k \leq \deg(f) + k \cdot \deg(t) \; .$$

In terms of d_0, we have

$$d_k \leq (k+1) \cdot d_0 \; .$$

The dependence on k is surely a weakness of the attack. We remark however again that for any of the three methods mentioned above, for the *generation* of some element a_{n+1}, the complete initial sequence $(a_1, \ldots, a_n)$ has to be computed and stored.

6 Outlook and comments

We have seen that expansion sequences should be regarded as cryptographically weak. It is now natural to define a new notion of complexity of sequences over $\mathbb{F}_q$:

For a finite sequence $\mathbf{a}$ of length n we define the *expansion complexity*, $E_{\mathbf{a}}$, as the minimum of the degrees of non-trivial polynomials $h(t,y) \in \mathbb{F}_q[t,y]$ with $h(t, \sum_{i=1}^{n} a_i t^i) = 0 \mod t^{n+1}$. Clearly, we always have $E_{\mathbf{a}} \leq L_{\mathbf{a}}$. We define the *expansion complexity profile* of a finite or infinite sequence $\mathbf{a}$ of length m as $(E_{\mathbf{a}}(n))_{n=1}^{m}$, where $E_{\mathbf{a}}(n)$ is the expansion complexity of the initial sequence of length n of $\mathbf{a}$. We remark that for a given sequence, one should not only consider the expansion complexity profile of the sequence but in fact all expansion complexity profiles for arbitrary starting points (that is, for the corresponding left-shifted sequences). In the realm of linear complexity, the corresponding suggestion was already made in [17], and a first study in this direction is [12].

The new notion of expansion complexity leads to some new research directions. One interesting task is to develop a probabilistic theory of expansion complexity, just as a probabilistic theory of linear complexity has been developed in [13] and [15]. A first task is here to study the probability distributions of expansion complexities for uniformly randomly distributed sequences of a fixed length.

After such a theory has been developed, pseudorandom generators for cipher streams should be analyzed from a statistic point of view via the theory. Let us note here that a corresponding statistic program for linear complexity is part of a test suite, issued by NIST, for pseudorandom sequences for cryptographic use ([7]).

Finally, we would like to make a remark on terminology: In [18], [14] and various other works, the authors speak of "sequences with perfect linear complexity profile". In the beginning of [20] the authors speak of sequences with "perfect linear complexity profile" or with "almost perfect linear complexity profile". However, later in this work and also in [16], the authors speak of "(almost) perfect" sequences instead. Moreover, sequences with d-almost perfect complexity profile are called "d-perfect". Here we would like to make the following remark: The previous expressions are completely adequate. However, the newer expressions are questionable. It is one thing to express that a mathematical object is "(almost) perfect" with respect to a particular aspect under consideration. It is however something else to say that a mathematical object is "(almost) perfect" by itself. Concretely, having in mind the results of this work, we would like to suggest to return to the older terminology and to discontinue speaking of "(almost) perfect sequences" and of "d-perfect sequences". Consequently, we have avoided

the usage of the expressions "(almost) perfect sequence" and "d-perfect sequence" in this article.

References

[1] M. Baker, E. González-Jiménez, J. González, and B. Poonen. Finiteness results for modular curves of genus at least 2. *Amer. J. Math.*, 127, 2005.

[2] W. Bosma, J. Cannon, C. Fieker, and A. Steel, editors. *Handbook of Magma functions, Edition 2.17.* 2011.

[3] H. Cohen. *A Course in Computational Algebraic Number Theory.* Springer-Verlag, 1996.

[4] H. Cohen. *Number Theory – Volume 1: Tools and Diophantine Equations.* Springer-Verlag, 2007.

[5] C. Diem. On arithmetic and the discrete logarithm problem in class groups of curves, 2008. Habilitation thesis.

[6] C. Diem. On the discrete logarithm problem in class groups of curves. *Math. Comp.*, 80:443 – 475, 2011.

[7] A. Rukhin et al. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publication 800-22, 2010.

[8] R. Hartshorne. *Algebraic Geometry.* Springer, 1977.

[9] F. Heß. Computing Riemann-Roch spaces in algebraic function fields and related topics. *J. Symb. Comput.*, 11, 2001.

[10] D. Kohel, S. Ling, and C. Xing. Explicit sequence expansions. In *Sequences and their Applications – Proceedings of SETA '98*, Discrete Mathematics and Theoretical Computer Science. Springer, 1999.

[11] S. Lang. *Algebraic Number Theory.* Springer-Verlag, 1994.

[12] H. Niederreiter. Keystream squences with a good linear complexity profile for every starting point. In J.-J. Quisquater and J. Vandewalle, editors, *Advances in Cryptology – Eurocrypt '89*, volume 434 of *LNCS*, pages 523 – 532. Springer-Verlag.

[13] H. Niederreiter. The probabilistic theory of linear complexity. In C. Günter, editor, *Advances in Cryptology – Eurocrypt '88*, volume 330 of *LNCS*, pages 191–209. Springer-Verlag, 1988.

[14] H. Niederreiter. Sequences with almost perfect linear complexity profile. In D. Chaum and W. Price, editors, *Advances in Cryptology – Eurocrypt '87*, volume 330 of *LNCS*, pages 37 – 51. Springer-Verlag, 1988.

[15] H. Niederreiter. A combinatorial approach to probabilistic results on the linear complexity profile ofrandom sequences. *J. Cryptology*, pages 105–112, 1990.

[16] H. Niederreiter and C. Xing. *Rational Points on Curves over Finite Fields*. Cambridge University Press, 2001.

[17] F. Piper. Stream ciphers. *Elektrotechnik und Maschinenbau*, 104:564 – 568, 1987.

[18] R. Rueppel. *Analysis and Design of Stream Ciphers*. Springer-Verlag, Berlin, 1986.

[19] A. Stein. Explicit infrastructure for real quadratic function fields and real hyperelliptic curves. *Grasnik Mathematicki*, 44:89 – 126, 2009.

[20] C. Xing and K. Lam. Sequences with almost perfect linear complexity profiles and curves over finite fields. *IEEE Trans. Infor. Theory*, pages 1267–1270, 1999.

[21] R. Zuccherato. The continued fraction algorithm and regulator for quadratic function fields of quaracteristic 2. *J. Algebra*, 563 – 587, 1997.

Claus Diem
University of Leipzig
Mathematical Institute
Johannisgasse 26
04103 Leipzig
Germany
diem@math.uni-leipzig.de