Observer Attack on Stream Ciphers

Ramachandran Anantharaman * **Virendra Sule** **

* *Department of Electrical Engineering, Indian Institute of Technology - Bombay, Mumbai, India (e-mail:ramachandran@ee.iitb.ac.in)*
** *Department of Electrical Engineering, Indian Institute of Technology - Bombay, Mumbai, India (e-mail:vrs@ee.iitb.ac.in)*

Abstract: This paper proposes an application of a new observer theory for non-linear systems developed previously to solve the Cryptanalysis problem of a special class of pseudorandom generators which are commonly used in Cryptography. The Cryptanalysis problem addressed here is that of the recovery of internal state of the non-linear dynamic stream generator from the output stream. The proposed methodology is termed as *observability attack*. It is also shown that for a special class of generators, the computations are of complexity $O(D^4)$ in pre-computation and of $O(D)$ for online computation, where $D = \sum_{i=0}^{d} \binom{n}{i}$ for this class of stream generators with n states and d the degree of the output function. The attack is technically applicable over general finite fields as well as most dynamic systems arising from models of stream ciphers and appropriate bounds on computation are estimated. From these complexity bounds, it follows that this attack is feasible in realistic cases and gives important estimates of time and memory resources required for Cryptanalysis of a class of stream ciphers.

Keywords: Stream ciphers, Cryptanalysis, Boolean Dynamical Systems, Observer Design, Koopman Operator

NOTATIONS AND PRELIMINARIES

$\mathbb{F}_q^n$ is the n-dimensional vector space over the finite field $\mathbb{F}_q$. The space V^o is the vector space of functions (including non-linear functions) from $\mathbb{F}_q^n \to \mathbb{F}_q$. The function $\chi_i(x) \in V^o$ is the i^{th}-coordinate function defined as $\chi_i(x) = x_i$. A monomial $\phi \in V^o$ is a function of the form $\phi(x) = \prod_i x_i^{d_i}$, where each $0 \leq d_i < q$.

1. NON-LINEAR STREAM GENERATORS

Filter generators and non-linear combiners are generic constructions used in stream ciphers and pseudorandom generators in Cryptography. Such generators can be modeled as a dynamical system over a finite field (DSFF) with state variables and outputs defined over $\mathbb{F}_q^n$ and $\mathbb{F}_q$, respectively. Such systems are designed to have one or more Feedback Shift Registers (FSRs) driving the update of internal states. In this paper, we call such pseudorandom generators, which depend on Linear FSRs (LFSRs) for internal state updates as *stream generators*. When the number of state variables is n, the state $x(k)$ of the system at any time k belongs to $\mathbb{F}_q^n$ and a *non-linear output map* $g : \mathbb{F}_q^n \to \mathbb{F}_q$ defines the output $z(k) = g(x(k))$. Typical constructions of such stream generators are shown in the Figures 1 and 2, with a single LFSR and multiple LFSRs, respectively, and a non-linear function g generating the output sequence. The internal state of the stream generator $x(k)$ is the collection of states of all registers of the LFSRs. Stream ciphers using such constructions are discussed in Rueppel (1986).

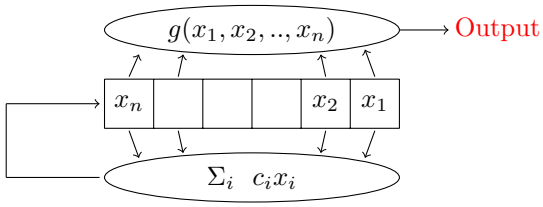

Fig. 1. Stream generator with with 1 LFSR

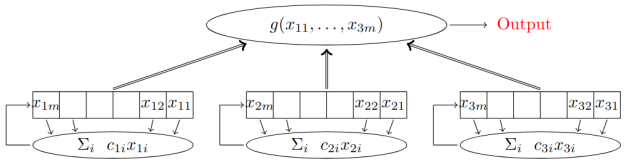

Fig. 2. Stream generator with 3 LFSRs

1.1 Cryptanalysis of stream generators

A major Cryptanalysis problem for such generators involves that of computing the initial condition of the generator $x(0)$ when an output stream $z(k)$ is made available over a limited length of time k, such as over an interval $[k_0, k_0 + m]$, $k_0 > 0, m > 0$. This problem is also known as *key recovery problem* of the stream generator, since the values of state variable at initial loading $x(0)$ consist of the symmetric key K (which is secret) and randomly chosen initializing values of states called IV (which is publicly known). Such a problem is NP for non-linear generators and is known to be computationally challenging as the number of states increases. Search for efficient algorithms for solving this Cryptanalysis problem of the generator

has continued ever since these have been found suitable for use in Cryptography. In this paper, we develop a new approach to the Cryptanalysis of non-linear stream generators called *Observability attack* based on the previous work Anantharaman and Sule (2021).

1.2 Previous work on Cryptanalysis of the stream generator

In the past, stream generators were Cryptanalyzed using correlation as well as algebraic attacks Klein (2013). In the former, correlation of the output stream $z(k)$ for $k \geq k_0$ with that of the internal states $x(k)$ is estimated. While correlation attack is statistical, the algebraic attack directly solves the non-linear polynomial system of equations with state variables as unknowns related to the output stream. Such a computation is NP and increases in complexity with the number of variables. Both attacks have not been known to scale up for realistic sizes of stream generators and work only when the number of internal states is small enough. The work in RØnjom and Helleseth (2007b); RØnjom et al. (2007); RØnjom and Helleseth (2007a) addressed the problem of internal state recovery of stream generators. The basic idea reported in these works is to construct a linear system model for the output stream in terms of monomials in the variables arising in the non-linear system of equations. A basis consisting of monomials is used to describe this linear system. This approach is broadly known as the extended linearization (XL) method of solving multivariate algebraic equations. The proposed method in our paper differs from the above by the method of construction of the linear model. Our approach constructs a restriction of the Koopman operator of the dynamic system representing the stream generator to the smallest invariant space containing the coordinate functions. Because of this minimal dimension of the invariant subspace, the dimension of the linear model constructed through our method will always be less than or equal to the dimension of the linear model constructed in RØnjom et al. (2007). Also, the computation of the internal state utilizes the well-known observer construction from systems theory. Hence our paper is an instance of applying the linear observer theory to non-linear dynamical systems over finite fields.

Observability of evolution of permutation maps [1] over a finite set X through a function f on X was discussed in the paper Byerly et al. (2003). However, this approach could not be utilized for the Cryptanalysis of stream ciphers because it assumes the base field for the function f to be the complex field. The state-space of the dynamics of the permutation under a complex field turns out to be an inner product space, and the permutation map action on functions on X is a normal operator. No such nice conditions hold when the field is finite. Hence the observability based approach to cryptanalysis of stream generators needed a fresh investigation after the paper Byerly et al. (2003). Another recent work on the observability of dynamical systems and its relevance to Cryptanalysis of stream ciphers is reported in Zhong and Lin (2016). This work utilizes what is known as the semi tensor product representation of Boolean functions and maps. While this paper is relevant to the problem posed here, we point out significant differences with our approach. First, the dimension of the linear model in the former approach is always exponential in n, and secondly, the approach is specifically only applicable to Boolean functions. The proposed approach in our paper is useful for realistic Cryptanalysis mainly because the dimension of the linear system obtained is not too large for the class of stream generators (and is never exponential). Moreover, our approach is applicable over any finite field and computationally feasible for fields with small characteristics.

[1] A permutation map is a bijection and most of the stream generators are designed to have an internal dynamics which is a bijection over $\mathbb{F}_q^n$

1.3 Contributions in this paper

This paper proposes a novel method for Cryptanalysis by developing an extension of the observer construction well known in systems theory to non-linear dynamical systems over finite fields. Termed as *Observer attack* to stream ciphers, the attack will work for any stream cipher with a linear or non-linear state update map, and with a linear or non-linear output function. Almost all stream ciphers fall into this category Rueppel (1986). The complexity analysis for such an attack is done for a special class of stream generators with a linear state update and non-linear output map, and it is shown that this attack is feasible in polynomial time for this special class. This type of construction of a linear dynamic observer for a non-linear dynamical system and its application to Cryptanalysis has never been known in previous literature. As an example, a stream generator of 80 bits with a non-linear output function is taken up for cryptanalysis, and its internal state is recomputed using the proposed attack by constructing a linear observer.

1.4 Mathematical model of the stream generator

Mathematically, any non-linear stream generator with linear state update and non-linear output map as in Figures 1 and 2 can be represented as a dynamical system in the following way

$$\begin{aligned} x(k+1) &= Ax(k) \\ z(k) &= g(x(k)) \end{aligned} \tag{1}$$

where $x(k) \in \mathbb{F}_q^n$ is the internal state, A an $n \times n$ matrix over $\mathbb{F}_q$ is the state-transition map, $g \in V^o$ is a non-linear function (the output function). When the stream generator is of the form 1, the A matrix is in companion form and the output is a non-linear function of the internal states. In the form 2, the A matrix is a block diagonal form representing the matrices in companion form of feedback polynomials of individual LFSRs and the output is a non-linear function g of all the states.

1.5 Koopman Linear System for Dynamical Systems over Finite Fields

Mathematically, any dynamical system over a finite field can be expressed as

$$\begin{aligned} x(k+1) &= F(x(k)) \\ z(k) &= g(x(k)) \end{aligned} \tag{2}$$

where $x(k) \in \mathbb{F}_q^n$, $z(k) \in \mathbb{F}_q^m$ are the internal state and outputs while $F : \mathbb{F}_q^n \to \mathbb{F}_q^n$, $g : \mathbb{F}_q^n \to \mathbb{F}_q^m$ are the state

transition and output map respectively. Let V^o be the vector space of $\mathbb{F}_q$-valued functions over $\mathbb{F}_q^n$. The Koopman operator F^* for the system (2) is a map from V^o to V^o defined as

$$F^*h(x) = (h \circ F)(x) = h(F(x)),$$

where $h \in V^o$ and $x \in \mathbb{F}_q^n$. The Koopman operator is linear over V^o, and the Koopman Linear System (KLS) corresponding to (2) is a linear dynamical system over V^o defined as

$$h_{k+1}(x) = F^*h_k(x)$$

for $h_k(x) \in V^o$. The paper by Anantharaman and Sule (2021) develops the theory of the Koopman operator for dynamical systems over finite fields and gives a linear algebraic formulation for computation of solution structures of a non-linear system using a reduction of the KLS. This system, called the Reduced Order Koopman Linear System (RO-KLS), is constructed by restricting the operator F^* to the smallest invariant subspace $W_1 \subseteq V^o$ consisting of the coordinate functions χ_i and the output functions $g_i(x)$. Necessary and sufficient conditions for observability of the original non-linear system are translated to the observability of the RO-KLS (which is a linear dynamical system). Also, a generic construction of dynamic observer was developed using the RO-KLS, and a unique reconstruction of the internal states of the non-linear system is possible whenever the RO-KLS is detectable. The following section describes the algorithm to compute the RO-KLS for stream generators.

2. RO-KLS FOR STREAM GENERATORS

The construction of the F^*-invariant subspace W_1 plays an integral part in constructing the RO-KLS. The dimension of the linear system is equal to the dimension of the smallest F^*-invariant subspace of V^o containing of the coordinate functions χ_i, $i = 1, \ldots, n$ and the output function g. The construction of this subspace is described in Algorithm 1.

Once the invariant subspace W_1 is computed, let its basis be $\mathcal{B} = \{\psi_1(x), \ldots, \psi_N(x)\}$. The RO-KLS (as evaluation map) is the linear system of dimension $dim(W_1)$ and can be expressed in terms of matrices with this specific basis $\mathcal{B}$. Let K_1^T be the restriction of Koopman operator on this invariant subspace W_1, C the matrix corresponding to the map from the basis functions $\mathcal{B}$ to the vector of coordinate functions $[\chi_1(x), \ldots, \chi_n(x)]^T$, and Γ be the matrix corresponding to representation of the function $g(x)$ in terms of the basis functions $\mathcal{B}$. Define the RO-KLS as follows

$$\begin{aligned} y(k+1) &= K_1 y(k) \\ w(k) &= Cy(k) \\ y_{op}(k) &= \Gamma y(k), \end{aligned} \tag{3}$$

where $y(k) \in \mathbb{F}_q^N$, $y_{op}(k) \in \mathbb{F}_q$ are the internal state and output of the RO-KLS respectively. Given any initial condition $x(0)$ of the non-linear stream generator and initiating the RO-KLS as

$$y(0) = \begin{bmatrix} \psi_1(x(0)) \\ \vdots \\ \psi_N(x(0)) \end{bmatrix},$$

Algorithm 1 Construction of W_1 - the smallest F^*-invariant subspace spanned by χ_i and g

procedure F^*-INVARIANT SUBSPACE(W_1)
  Outputs:
  W_1 - the smallest F^*-invariant subspace containing the coordinate functions χ_i and the non-linear function g.
  $\mathcal{B}$ - the basis for the invariant subspace W_1
  Compute the cyclic Subspace
    $Z(\chi_1; F^*) = \langle \chi_1, F^*\chi_1, \ldots, (F^*)^{l_1-1}\chi_1 \rangle$
  Set of basis functions
    $\mathcal{B} = \{\chi_1, F^*\chi_1, \ldots, (F^*)^{l_1-1}\chi_1\}$
  if $\chi_2, \chi_3, \ldots, \chi_n \in \text{Span}(\mathcal{B})$ **then**
    $W_1 \leftarrow \text{Span}(\mathcal{B})$
    go to 14
  else
    Find the smallest i such that $\chi_i \notin \text{span}(\mathcal{B})$
    Compute the smallest l_i such that
    $(F^*)^{l_i}\chi_i \in \text{Span}\{\mathcal{B} \cup \langle \chi_i, F^*\chi_i, \ldots, (F^*)^{l_i-1}\chi_i \rangle\}$
    $V_i = \{\chi_i, F^*\chi_i, \ldots, (F^*)^{l_i-1}\chi_i\}$
    Append the set V_i to $\mathcal{B}$
    go to 5
  if $g \in \text{Span}(\mathcal{B})$ **then**
    halt
  else
    Compute the smallest j such that
      $(F^*)^j g \in \text{Span}(\mathcal{B} \cup \langle g, F^*g, \ldots, (F^*)^{j-1}g \rangle)$
    $V_g = \{g, F^*g, \ldots, (F^*)^{j-1}g\}$
    Append the set V_g to $\mathcal{B}$
    halt

it is proved that the output sequence $y_{op}(k)$ of the RO-KLS is the same as the output sequence of the stream generator (1) initiated with the same $x(0)$.

2.1 Dimension of W_1

Given the RO-KLS as in (3), the first question which needs to be answered is "*Is there any bound on the dimension of W_1?*". The stream generator as in (1) is one of few systems for which this question can be answered convincingly in the affirmative. Since the internal dynamics of the stream generator is linear, the dimension of the RO-KLS solely depends on the non-linear output function g.

Lemma 1. Given a dynamical system over a finite field with linear internal dynamics as in (1) and a monomial ϕ, then

$$\text{degree } (F^*\phi) \ \leq \ \text{degree } (\phi),$$

where F^* is the Koopman operator.

Proof. Assume that the system (1) evolves over $\mathbb{F}_q^n$, where $q = p^m$. Let the monomial be

$$\phi(x_1, \ldots, x_n) = \prod_{j=1}^{n} x_j^{d_j},$$

where $0 \leq d_j < q-1$ and the degree of the monomial ϕ is $\sum_j d_j$. The action $F^*\phi(x_1, \ldots, x_n)$ is defined as

$$F^*\phi(x_1, .., x_n) = \phi(f_1(x_1, .., x_n), .., f_n(x_1, .., x_n)),$$

where $f_1, \ldots, f_n$ are functions corresponding to the state transition for each x_i. Since the internal dynamics of the stream generator is linear, each function f_i can be written as

$$f_i(x_1, \ldots, x_n) = \sum a_{ij} x_j,$$

and these a_{ij} are entries of the matrix A in (1). In particular,

$$F^*\phi = F^*(\prod_j x_j^{d_j}) = \prod_j (\sum_{k=1}^{n} a_{jk} x_k)^{d_j}.$$

This means

$$\text{degree } (F^*\phi) = \text{degree } \left(\prod_j (\sum_{k=1}^{n} a_{jk} x_k)^{d_j} \right).$$

Also,

$$\text{degree } \left((\sum_{k=1}^{n} a_{jk} x_k)^{d_j} \right) \leq d_j.$$

The less-than sign is because there can be a case where all a_{jk} can be zero. So, each term in the product

$$\text{degree } \left(\prod_j (\sum_{k=1}^{n} a_{jk} x_k)^{d_j} \right)$$

has a degree $\leq d_j$, and hence

$$\begin{aligned} \text{degree } (F^*\phi) &= \text{degree } \left(\prod_j (\sum_{k=1}^{n} a_{jk} x_k)^{d_j} \right) \\ &\leq \sum_j d_j = \text{degree } (\phi). \qquad \square \end{aligned}$$

Remark 2.1. Given any non-linear function g, it can be written as a sum of monomials and the function g has a degree d_g, the largest degree of the constituent monomials. Since Koopman operator is linear over V^o and using the above lemma, it can be seen that the action of F^* on g does not increase the degree of g. Hence, for a stream generator (1)

$$\text{degree}(F^*g) \leq \text{degree}(g) \quad \forall\ g \in V^o.$$

The following theorem gives the upper bound on the dimension of W_1

Theorem 2. Given a non-linear stream generator as in (1) over $\mathbb{F}_q^n$ with the output g having a degree d, the dimension of the F^*-invariant subspace W_1 is bounded by

$$\frac{(1 - n^{d+1})}{1 - n} \approx O(n^d).$$

Proof. Since the degree of g is d, and from lemma (1), the invariant subspace W_1 can have functions only upto degree d. A counting of all the independent monomials over n-variables from $\mathbb{F}_q^n \to \mathbb{F}_q$ gives an upper bound on the dimension of the invariant subspace W_1.

For example, with degree r, the total number of independent monomials is upper bounded by n^r since the degree is r and there are n variables to choose from it and the variables can get repeated too (which means that in the specific monomial, that variable has a power > 1). Hence, the conservative estimate on the number of independent monomials of degree r over n variables is n^r.

So, a function g having degree d can have other terms of degree d or less. So, counting all the independent monomials of degree d or less gives an upper bound on the dimension of W_1. This gives

$$\dim\ (W_1) \leq 1 + n + n^2 + \cdots + n^d,$$

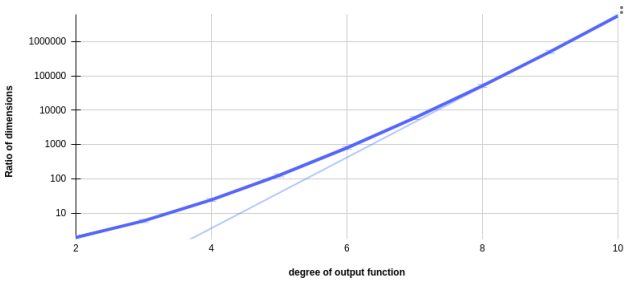

Fig. 3. Ratio of $n^d/dim(W_1)$ vs the degree d for an 80-bit stream generator

where 1 is for the constant function, n is for linear functions and so on. The expression is the partial sum of the geometric series and simplifies to $\frac{1-n^{d+1}}{1-n}$. $\square$

Remark 2.2. Note that the estimate in Theorem 2 does not take the field equation [2] into account. For $d \geq q$, the powers x_i^r, $(r > q)$ in the monomial gets reduced to a power $x_i^{r_d}$ where $r_d = r \bmod q$ and one can have better estimates to the dimension of W_1 for specific fields. Also, the estimate in Theorem 2 is the best upper bound on the dimension of W_1 whenever $d < q$.

When the stream generator is over $\mathbb{F}_2$, any variable x_i satisfies the equation $x_i^2 = x_i$ (as functions) which drastically reduces the dimension of W_1 and leads to the following corollary.

Corollary 2.1. Given a stream generator as in (1) with $x(k)$ over $\mathbb{F}_2^n$, with the non-linear function having degree d, the dimension of W_1 is upper bounded by

$$\dim\ (W_1) \leq \sum_{i=0}^{d} \binom{n}{i}.$$

This can be proved by a simple counting argument of the number of distinct monomials of degree less than or equal to d over n variables. It is pertinent to see that when the degree d is small, the dimension of the subspace W_1 is much smaller when compared with the dimension of the space V^o (which is of exponential size 2^n, for an n-state stream generator).

Table 1 compares the upper bounds for any 80-bit stream generator over $\mathbb{F}_2$ for small degrees (d) of the non-linear function g. The maximum dimension of the subspace W_1 are computed for each d using Corollary 2.1. Although Theorem 2 and Corollary 2.1 give a polynomial upper bound on the dimension of W_1, it can be seen from Table 1 that Corollary 2.1 provides a refined bound for the dimension of W_1. Further, as a comparison, the ratio of the upper bound on the dimension W_1 to the dimension of the space V_0 is also given.

Remark 2.3. Figure 3 shows the ratio of n^d to $dim(W_1)$ for different degrees d of the output function for an 80-bit stream generator over $\mathbb{F}_2$. It can be seen that the graph (dark blue line) grows approximately linear in the log scale. Hence it can be concluded that the dimension of W_1 for the stream generator over $\mathbb{F}_2$ is much smaller than n^d. The light blue line in the figure shows the best exponential approximation of the data points.

[2] For $\mathbb{F}_q$, the field equation is $x^q - x = 0$

Table 1: Upper bounds on dimension of W_1 for different degrees of output function for a 80 variable stream generator over $\mathbb{F}_2$

degree (d)	*maximum dim of* W_1	n^d	n^d/*max dim* W_1	*max dim*(W_1)/*dim*(Vo)
1	81	80	1	6.7×10^{-23}
2	3241	6400	1.97	2.7×10^{-21}
3	8.54×10^4	5.12×10^5	6	7×10^{-20}
4	1.67×10^6	4.1×10^7	24.57	1.4×10^{-18}
5	2.57×10^7	3.28×10^9	127.47	2.1×10^{-17}
6	3.26×10^8	2.62×10^{11}	803.6	2.7×10^{-16}

The following section focuses on the construction of the observer using the RO-KLS for the stream generator.

3. COMPUTATION OF INTERNAL STATE FOR A STREAM GENERATOR OVER $\mathbb{F}_2^N$

Given the linear system (3) starting from an internal state $y(k_0)$, the output $y_{op}(k)$ at each $k \geq k_0$ is given by

$$y_{op}(k_0 + k) = \Gamma y(k_0 + k) = \Gamma \ K_1^k y(k_0).$$

Given the output $z(k), \quad k = k_0, k_0 + 1, \ldots,$ generated by the non-linear stream generator (1) starting from an initial condition $x(k_0)$, it is proved in Anantharaman and Sule (2021) that when the RO-KLS (3) is initiated with $y(k_0)$ as

$$y(k_0) = \begin{bmatrix} \psi_1(x(k_0)) \\ \psi_2(x(k_0)) \\ \vdots \\ \psi_N(x(k_0)) \end{bmatrix},$$

where $\psi_i(x),\ i = 1, \ldots, N$ form the basis of $\mathcal{B}$, then the output sequence $y_{op}(k), \quad k \geq k_0, \ldots$ generated by (3) is the same as $z(k)$, the output of the non-linear stream generator (1). Since RO-KLS is a linear system, it follows that

$$\begin{bmatrix} y_{op}(k_0) \\ y_{op}(k_0 + 1) \\ \vdots \\ y_{op}(k_0 + N) \end{bmatrix} = \begin{bmatrix} \Gamma \\ K_1\Gamma \\ \vdots \\ K_1^{N-1}\Gamma \end{bmatrix} y(k_0) =: \mathcal{O}\ y(k_0), \qquad (4)$$

where $\mathcal{O}$ is the *Observability matrix* corresponding to the linear system (3) with state matrix K_1 and output matrix Γ. To retrieve $x(k_0)$ given the sequence of outputs $z(k),\ k \geq k_0$ of the stream generator, the linear system of equations (4) needs to be solved for $y(k_0)$ with $y_{op}(k) = z(k)$ and then compute $x(k_0)$ as $x(k_0) = Cy(k_0)$ using C defined in (3). A unique solution for (4) exists if the observability matrix $\mathcal{O}$ is of full rank or equivalently if the system (3) is *Observable*. When the matrix $\mathcal{O}$ is not of full rank, then multiple solutions $y(k_0)$ exist for the given stream of outputs. This leads to multiple $x(k_0)$ through the map C.

3.1 Dynamic Observer

In the previous part, the RO-KLS is constructed for the stream generator, and the internal states $x(k_0)$ corresponding to the sequence of outputs $z(k),\ k \geq k_0$ could be computed using linear algebraic computations. Furthermore, in this section, we construct a *dynamic observer*, which takes the output of the non-linear stream generator $z(k)$ as an input and compute the internal state $x(k)$ of the non-linear stream generator. Mathematically the dynamic observer is a dynamical system defined as

$$\begin{aligned} \hat{y}(k+1) &= K_1\hat{y}(k) + L(z(k) - y_{op}(k)) \\ \hat{x}(k) &= C\hat{y}(k), \end{aligned} \qquad (5)$$

where $\hat{y}(k) \in \mathbb{F}_q^n$ is the observer state, $z(k)$ is the output of the stream generator, $\hat{x}(k)$ is the computed internal state of the stream generator, K_1 and C are as defined as in (3). The matrix L, known as *observer gain* is chosen such that $K_1 - L\Gamma$ is nilpotent.

Remark 3.1. Choosing L such that $K_1 - L\Gamma$ is nilpotent will make the observer error go to zero in finite time instants. Also, when the linear system is observable, arbitrary assignment of the characteristic polynomial of $K_1 - L\Gamma$ is possible, and hence all the characteristic polynomial is chosen to be x^N, which makes $K_1 - L\Gamma$ nilpotent.

From the linear systems theory, it is known that whenever the system (3) is observable, such an L always exists. Also, the notion of detectability of a linear dynamical system over finite fields is defined in Anantharaman and Sule (2021), which is reproduced here. Whenever L exists such that $(K_1 - L\Gamma)$ is nilpotent, the system (3) is defined to be *detectable*. The set of observable linear systems is a subset of detectable linear systems.

Given the stream generator (1) and its RO-KLS as constructed in (3), the dynamic observer construction is graphically illustrated in Figure 4. This uniqueness of the construction is that the observer has linear internal dynamics and reconstructs the internal state of a non-linear dynamical system.

For an available output sequence starting from time k_0, the observer states can be initialized to any arbitrary initial condition $\hat{y}(k_0)$ and whenever the RO-KLS is detectable, the computed internal state of the stream generator $\hat{x}(k)$ converges to the true internal state of the stream generator in maximum N_0 time instants, where N_0 is the index of nilpotence of $(K_1 - L\Gamma)$.

4. SOLUTION OF THE KEY RECOVERY PROBLEM FROM AN ARBITRARY INTERNAL STATE $X(K)$ - THE OBSERVABILITY ATTACK

In this section, we describe the complete algorithm for recovery of initial condition $x(0)$, also known as the key recovery problem, as the secret key used in the encryption is either the entire $x(0)$ or a part of it using the RO-KLS for the stream generator. Given a sequence of outputs of the stream generator $z(k)$ starting from k_0, it can be seen from the previous section that whenever the corresponding RO-KLS of the stream generator is observable (the $\mathcal{O}$ being full rank), the internal state $x(k_0)$ can be computed uniquely. Under the assumption that the system is detectable, the internal state of the stream generator can be uniquely computed at a time instant $k_0 + N_0$, where N_0 is the index of nilpotence of $K_1 - L\Gamma$.

Once this internal state is uniquely computed at some $x(k)$, the initial condition $x(0)$ is computed by revers-

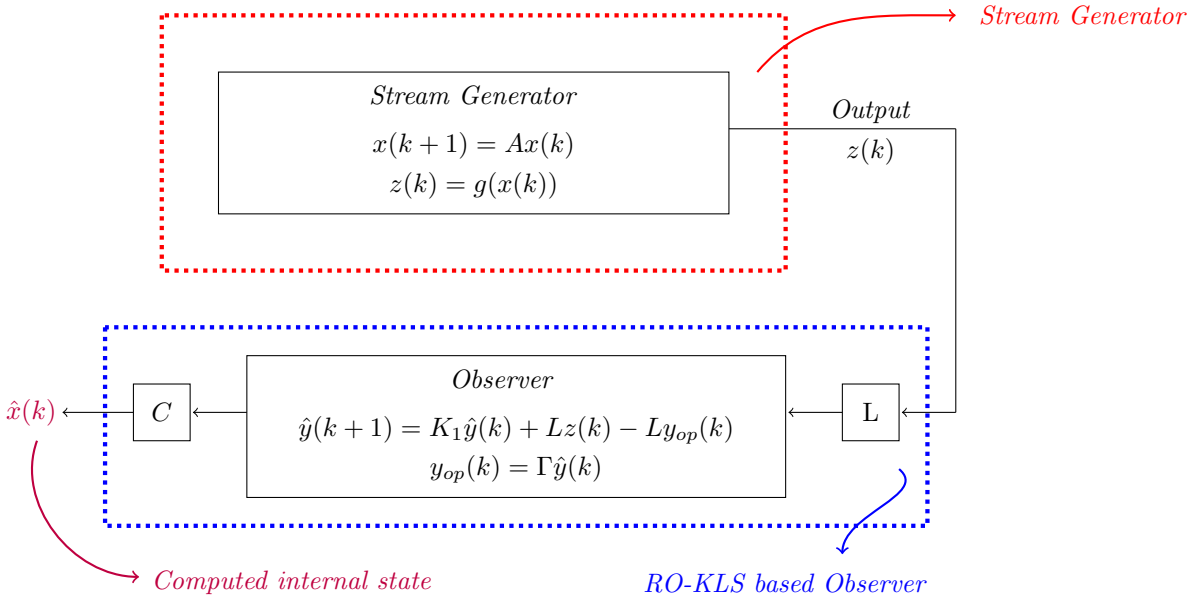

Fig. 4. Dynamic Observer for Stream generator using RO-KLS

ing the dynamics of the stream generator. However, to uniquely reverse the dynamics, the internal dynamics of the stream generator should be reversible, or equivalently the matrix A needs to be non-singular. Whenever the internal dynamics of a system is an $1-1$ map over $\mathbb{F}_q^n$, the dynamical system is said to be *non-singular*. It is easy to see that whenever the system is non-singular and RO-KLS being detectable, unique retrieval of $x(0)$ is possible from the output sequences starting from any time instant k_0.

If the internal dynamics of the stream generator is not a $1-1$ map over $\mathbb{F}_q^n$, but the RO-KLS is detectable, then instead of a unique $x(0)$, there would be a family of initial conditions corresponding to the unique $x(k_0+N_0)$ reconstructed from the dynamic observer as solutions of the following equation

$$A^{k_0+N_0}x(0) = x(k_0+N_0)$$

When the RO-KLS is neither detectable nor observable, then there are multiple initial conditions $y(k_0)$ for the output sequence $z(k)$ which can be computed through equation (4). Corresponding to these $y(k_0)$ solutions, there exists multiple points $x(k_0)$ in the state space of the stream generator. These correspond to possibly multiple symmetric keys in the initial state $x(0)$. In practice, however, superfluous multiple keys corresponding to the same output stream rarely exist as these denote redundant keys. Stream ciphers are almost never designed to have such redundancies. Hence stream generators are rarely likely to be unobservable.

Algorithm 2 summarizes the discussion about the retrieval of initial condition $x(0)$ from the sequence of outputs $z(k_0), z(k_0+1), \ldots$, using the RO-KLS.

5. EXAMPLE: AN 80-BIT STREAM GENERATOR

Consider the 80-bit stream generator made up of a single LFSR of 80 bit with the characteristic polynomial $p(x)$ as

$$p(x) = x^{80} + x^{53} + x^{47} + x^{35} + x^{33} + x^{10} + 1.$$

The characteristic polynomial determines the feedback coefficients of the LFSR. The internal dynamics is given by a matrix in companion form with its characteristic polynomial $p(x)$. The non-linear output function is considered to be a majority function of three internal states chosen as below.

$$\begin{aligned} g(x_1, \ldots, x_{80}) &= \text{Majority}(x_1, x_{26}, x_{52}) \\ &= x_1x_{26} + x_1x_{52} + x_{26}x_{52}. \end{aligned}$$

The dimension of the subspace W_1 is computed to be 3240. The RO-KLS is a linear system of dimension 3240 with the matrices $K_1 \in \mathbb{F}_2^{3240\times3240}$, $\Gamma \in \mathbb{F}_2^{1\times3240}$ and $C \in \mathbb{F}_2^{80\times3240}$ and

$$\begin{aligned} y(k+1) &= K_1y(k) \\ y_{op}(k) &= \Gamma y(k) \\ x(k) &= Cy(k). \end{aligned}$$

The RO-KLS is verified to be *observable* and hence there exists a matrix $L \in \mathbb{F}_2^{3240\times1}$ such that $K_1 - L\Gamma$ is a nilpotent matrix. The internal state of the observer is $\hat{y}(k)$ and its dynamics

$$\hat{y}(k+1) = K_1\hat{y}(k) + Lz(k).$$

The reconstructed internal state of the stream generator is $\hat{x}(k)$, and computed as

$$\hat{x}(k) = C\hat{y}(k).$$

The observer is initiated with random $\hat{y}(0)$ and it can be seen that the computed state $\hat{x}(k)$ converges to the internal state $x(k)$ of the stream generator within a 3240 time instances, which is the dimension of the RO-KLS.

To verify the correctness of the observer, let the difference in estimation at each time instant be $e(k) = x(k) - \hat{x}(k)$. For convenience, the Hamming distance[3] of $e(k)$ is chosen as a metric. It is seen from Figure 5 that the Hamming distance of the error is continuously zero after 3240 time instances, indicating that the internal state is reconstructed exactly.

[3] The Hamming distance for a vector over $\mathbb{F}_2$ is the number of non-zero entries in that vector.

Algorithm 2 Retrieval of initial condition $x(0)$ using observerability attack

procedure OBSERVABILITY ATTACK
  Outputs:
  - Reconstruction of internal state $x(k_0 + l)$ from the output sequence $z(k)$ starting from k_0
  - Retrieval of initial condition $x(0)$ of the stream generator.

  Compute the invariant subspace W_1 using the Algorithm 1 and then construct the RO-KLS of the stream generator.
  if RO-KLS detectable **then**
    Construct the dynamic observer as in Figure 4 and reconstruct the internal state uniquely at $x(k_0 + l)$, l is the index of nilpotence of $K_1 + L\Gamma$
    if Internal dynamics of stream generator reversible **then**
      Compute the initial condition $x(0)$ uniquely from the unique $x(k_0 + l)$
    else
      Compute the all possible initial condition $x(0)$ satisfying $A^{(k_0+l)}x(0) = x(k_0 + l)$
  else
    Solve for all the solutions $y(k_0)$ of the linear equation (4) for the given output sequence.
    The internal states $x(k_0) = Cy(k_0)$ are the set of possible states which could generate the output sequence.
    Compute the solution(s) of $A^{k_0}x(0) = x(k_0)$
  halt

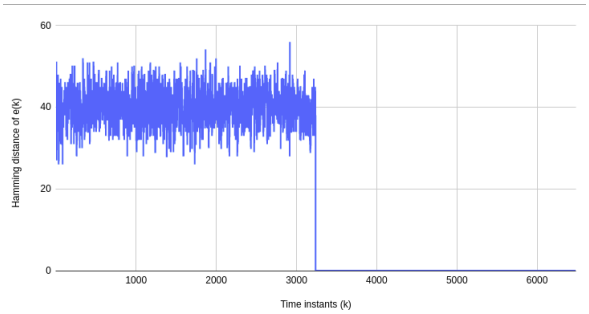

Fig. 5. Hamming distance of the error between the internal state of the stream generator and the reproduced state through dynamic observer

6. COMPUTATIONAL COMPLEXITY OF COMPUTING INTERNAL STATES

The final part of this paper is focused on the analysis of computational complexity of the proposed observability attack. The complexity of computing the internal states of the filter generator is primarily dependent on the dimension of the RO-KLS. Once the RO-KLS is computed, further complexities are polynomial in the dimension of the RO-KLS. The overall computations for recovery of the internal state of a stream generator can be divided into two parts. The first part deals with the construction of the RO-KLS, which is offline (and need to be done once for a given stream generator), and the second (the online part) is the recovery of the internal state of the filter generator from the given output stream $z(k)$ using the RO-KLS.

6.1 Preliminary offline computations

The offline computation concerns with the construction of the RO-KLS from a given non-linear filter generator. From Theorem 2, the dimension of the RO-KLS depends on the degree of the output function $g(x)$. Let D be the maximum possible dimension of this subspace (which is equal to the number of independent functions in n-variables with degree less than or equal to d and D is given as in Theorem 2 or Corollary 2.1 depending on the field.). Let $\mathcal{S}$ be a space of functions over $\mathbb{F}_q^n$ of degree less than or equal to d. So any function of degree less than or equal to d can be written as a linear combination of a chosen basis of $\mathcal{S}$ and hence represented as a vector in $\mathbb{F}_q^D$. For example, in a 4 bit filter generator over $\mathbb{F}_2$ with the output restricted to degree 2, one ordered-basis for $\mathcal{S}$ is given below

$$\mathcal{B}_\mathcal{S} = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}$$

where 1 is the constant function. Such a basis is referred to as the *monomial basis*. For example, the function

$$h(x_1, x_2, x_3, x_4) = x_1 + x_1x_3 + x_2x_4,$$

can be represented as the vector $[0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0]^T$ with the monomial basis $\mathcal{B}_\mathcal{S}$. Computation of a basis for W_1, the invariant subspace spanning the coordinate and output functions is the main part of the RO-KLS construction. Let the output function g be represented as vector $v_g \in \mathbb{F}_q^D$ with a chosen basis of $\mathcal{S}$. Let $g_i = (F^*)^i g$ denote the action of Koopman operator F^* i-times on the function g. From Theorem 2, it is known that deg $F^*g \leq d$, (d = degree of g) and hence every iterate $(F^*)^i$ on g is of degree $\leq d$ and hence in the span of $\mathcal{S}$. Each of these iterates $(F^*)^i$ g can be represented as a vector v_{g_i} over $\mathbb{F}_q^D$. Similarly, all the coordinate functions χ_i are in the span of $\mathcal{S}$ and hence have an unique representation as a vector v_{x_i}. Starting with v_g, one needs to find the smallest l such that the vector v_{g_l} is linearly dependent on

$$v_{x_1}, \ldots, v_{x_n}, v_g, v_{g_1}, \ldots, v_{g_{l-1}}. \tag{6}$$

This is a linear algebraic computation over the vectors $\mathbb{F}_q^D$, and is of order D^3. In the worst case, the dimension of W_1 is going to be D, and hence D such linear dependencies are to be checked. Thus the total offline computations will be of the order D^4.

6.2 Online computations

Once the linear model of the filter generator is computed, the dynamic observer does only forward computations of the evolution of the RO-KLS. Let N be the dimension of W_1. At each stage, the observer (5) updates the internal state as

$$\hat{y}(k + 1) = K_1\hat{y}(k) + L(z(k) - y_{op}(k))$$

The second part of the update $L(z(k)-y_{op}(k))$ is a vector-scalar multiplication as both $z(k)$ and $y_{op}(k)$ are scalars and L is a vector of length N and is of complexity $O(N)$. The first part $K_1\hat{y}(k)$ is matrix-vector multiplication. In general, it is of order $O(N^2)$, but we look to exploit the structure of K_1. Choosing the basis of W_1 as in equation (6), it can be seen that the matrix representation of K_1 is a block triangular matrix.

$$K_1 = \begin{bmatrix} K_{11} & 0 \\ K_{21} & K_{22} \end{bmatrix}$$

where K_{11} is the system matrix A as in equation (1). K_{21} and K_{22} are defined as

$$K_{21} = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \\ \alpha_1 & \dots & \alpha_n \end{bmatrix} \quad K_{22} = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 1 \\ \beta_0 & \beta_1 & \dots & \beta_{l-1} \end{bmatrix}$$

where α_i and β_i are from the linear dependency relation

$$(f^*)^l(x) = \sum_{i=1}^{n} \alpha_i \chi_i(x) + \sum_{i=1}^{l-1} \beta_i (f^*)^i(x)$$

Using this specific choice of $\mathcal{B}$ and the above decomposition of the matrix K_1, the computations become simpler. Let $\hat{y}(k) = [\hat{y}_1(k) \;\; \hat{y}_2(k)]^T$ where the dimension of $\hat{y}_1$ is n and $\hat{y}_2$ is l where, $l = N - n$. So the computation of $K_1\hat{y}(k)$ is equal to computing $K_{11}\hat{y}_1(k)$ and $K_{21}\hat{y}_1(k)$ and $K_{22}\hat{y}_2(k)$. The computation efforts required for these are explained below.

- $K_{11}\hat{y}_1(k)$ is a matrix-vector multiplication of dimension n. Assuming no structure of A, the total number of operations is n^2.
- $K_{21}\hat{y}_1(k)$ is a matrix-vector multiplication. The dimension of K_{21} is $l \times n$. But the first $l-1$ rows of K_{21} are 0 and hence it is only a vector-vector product with total n operations.
- K_{22} is in companion form. The first $l-1$ rows of the product requires only the computation of $K_{22}[i, i+1]\hat{y}_2[i+1]$ each of which is $O(1)$ and hence, a cumulative of $l-1$ operations. The l^{th}-row is a vector-vector product of dimension l and a total of l operations. Hence the total number of operations is $2l-1$.

Cumulatively, there is a total of $n^2+n+2l-1$ operations. We know that $l = D - n$ and $l >> n$. Hence the total complexity of online computations is $O(N)$. Also, the computation $\hat{x}(k) = C\hat{y}(k)$ is needed to compute the internal state of the filter generator. There are totally $n^2 l$ operations. And with $l >> n$ the computations are of order $O(N)$. Hence, the effective online computations are of order $O(N)$.

Also, the reconstructed internal state through the observer converges to the internal state of the filter in M time instants where M is the index of nilpotence of $K_1 - L\Gamma$. The total online computations to reconstruct the internal state are order $O(NM)$.

Remark 6.1. It is to be noted that the dimension N of the subspace W_1 has an upper bound D. So, in essence, the online computations are of order $O(D)$.

7. CONCLUSION

This paper proposes a new methodology for Cryptanalysis of stream ciphers. An observer for a non-linear dynamical system over a finite field is designed using the Koopman operator to reconstruct the internal state of the non-linear dynamical system. This construction uses only linear algebraic computations. Termed as Observability attack, this approach is shown to reconstruct the internal state of a special class of stream generators in polynomial time on the dimension of the internal state of the stream generator. Though this paper focuses on a special class, this approach for reconstructing the internal state is generic and can be applied to any generic pseudorandom generator with linear or non-linear state update and output function.

REFERENCES

Anantharaman, R. and Sule, V. (2021). Koopman operator approach for computing structure of solutions and observability of nonlinear dynamical systems over finite fields. *Math. Control Signals Syst.*, 33, 331–358.

Byerly, R.E., Drager, L.D., and Lee, J.M. (2003). Observability of permutations, and stream ciphers. *IEEE Transactions on Information Theory*, 49, 3326–3330.

Klein, A. (2013). *Stream Ciphers*. Springer-Verlag.

Rueppel, R.A. (1986). *Analysis and Design of Stream Ciphers*. Springer Berlin Heidelberg.

RØnjom, S., Gong, G., and Helleseth, T. (2007). On attacks on filtering generators using linear subspace structures. *Proceedings of International Workshop - Sequences, Subsequences and Consequences*, 204–217.

RØnjom, S. and Helleseth, T. (2007a). The linear vector space spanned by the nonlinear filter generator. *Proceedings of International Workshop - Sequences, Subsequences and Consequences*, 169–183.

RØnjom, S. and Helleseth, T. (2007b). A new attack on filter generator. *IEEE Transactions on Information Theory*, 53, 1752–1757.

Zhong, J. and Lin, D. (2016). Linearization of nonlinear filter generators and application to cryptanalysis of stream ciphers. *Journal of Complexity*, 35, 29–45.