Message-Embedded Cryptosystems: Cryptanalysis and Identifiability
Floriane Anstett, Gilles Millérioux, Gérard Bloch

To cite this version:

HAL Id: hal-00114178
https://hal.archives-ouvertes.fr/hal-00114178
Submitted on 15 Nov 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Message-Embedded Cryptosystems: Cryptanalysis and Identifiability

Floriane Anstett, Gilles Millerioux and Gérard Bloch*

Abstract

The aim of this paper is to compare two encryption schemes, the standard stream cipher and a so-called message-embedded cryptosystem. The comparison is based on two main aspects. The first aspect deals with the synchronization of the time-varying keys at the transmission side and at the reception side. The second aspect focuses on the cryptanalysis of the encryption algorithms. The cryptanalysis focuses on the system parameter retrieving. For message-embedded cryptosystems, the cryptanalysis is thus treated as a parametric identifiability issue. Two methods, the local state isomorphism approach and the Gröbner bases method, are presented for systems including polynomial nonlinearities. It is shown that these systems are weak against algebraic attack.

1 INTRODUCTION

Since 1993, a lot of methods involving nonlinear dynamic systems in order to mask an information have been proposed, because these systems can exhibit complex behaviors. In particular, the chaotic behaviors can be distinguished by their extreme sensitivity to initial conditions and to model parameters changes. Thus, the signals resulting from chaotic systems are broadband, long-term unpredictable and present random-like statistical properties although they are generated by deterministic systems. That is why, there is likely a connection between the random look behaviors exhibited by chaotic systems and the required properties like confusion and diffusion of cryptosystems. The chaotic masking [1], the parametric modulation [2], the approach by inclusion [3] (and references therein) have been proposed. An overview of these different methods can be found in [4]. However, very few works have really established the connection between the standard encryption algorithms and those based on the generation of chaotic sequences, except for deep comparative studies in [5][6]. In particular, the cryptanalysis of the chaos-based encryption algorithms is really missing today, although it constitutes an essential step of their validations.

The aim of this paper is to compare two encryption schemes, the standard stream cipher and a so-called message-embedded cryptosystem. The comparison is based on the synchronization of the time-varying keys, also called running keys, at the transmission and at the reception and on the cryptanalysis of the encryption algorithms, studied through the particular problem of the static key reconstruction. In the case of the message-embedded cryptosystem, the cryptanalysis is treated here as a parametric identifiability problem, borrowed from the control theory. Note that the parametric identifiability has been evoked and illustrated for the first time in [7], but, until now, has never been really formalized. The paper is organized as follows. In Section 2.1, we recall the principle of usual stream cipher. Then, in Section 2.2, we present the message-embedded cryptosystem, its principle and the information reconstruction based on observers. Finally, the Section 3 deals with the static key reconstruction problem and with the parametric identifiability. The local state isomorphism approach and the Gröbner bases method are presented for testing the parametric identifiability for the suitable choice of the static key, in the case of systems including polynomial nonlinearities. In Section 4, two examples emphasize the weakness of such systems.

*The corresponding authors are with Centre de Recherche en Automatique de Nancy, Université Henri Poincaré - Nancy 1, CRAN - ESSTIN, 2 Rue Jean Lamour 54519 Vandoeuvre Les Nancy Cedex, France. Email : floriane.anstett@essttn.uhp-nancy.fr, gilles.millerioux@essttn.uhp-nancy.fr, gerard.bloch@essttn.uhp-nancy.fr
2 Principles of Encryption: Comparison

2.1 Usual Stream Cipher

There exists two common classes of stream cipher, one is called synchronous and the other self-synchronous [8]. They are respectively illustrated on the Figures 1(a) and 1(b).

![Stream cipher: (a) synchronous, (b) self-synchronous](image)

Figure 1: Stream cipher: (a) synchronous, (b) self-synchronous

2.1.1 Transmitter and encryption

The synchronous stream cipher obeys, at the transmitter side:

\[
\begin{cases}
K_k = f_{\theta}(K_{k-1}) \\
c_k = e(K_k, m_k)
\end{cases}
\]

(1)

In this encryption scheme, the information signal, also called the plaintext, is divided into blocks of same length, called symbols and denoted by \(m_k\). The encryption transformation \(e\) can change for each symbol since \(e\) depends on a time-varying key \(K_k\) which is called keystream. The keystream \(K_k\) is generated by a function \(f_{\theta}\), parameterized by a constant \(\theta\). Usually, the plaintext \(m_k\) and the ciphertext \(c_k\) are binary words, the function \(e\) is a simple XOR operation. If the running key \(K_k\) is randomly chosen and never used again, the encryption scheme is called “one-time pad”.

Generally, the running key is generated iteratively by feedback shift registers which produce pseudo-random sequences, as, for instance, the Linear Feedback Shift Registers (LFSR). The ciphertext \(c_k\) is available at the transmitter output.

The self-synchronous stream cipher obeys, at the transmitter side:

\[
\begin{cases}
K_k = f_{\theta}(c_{k-1}, \ldots, c_{k-t}) \\
c_k = e(K_k, m_k)
\end{cases}
\]

(2)

\(f_{\theta}\) is a function parameterized by the constant parameter \(\theta\), which generates the keystream \(K_k\). Unlike the synchronous stream cipher, \(K_k\) does not depend on an internal dynamic but only on a fixed number of past values of \(c_k\). However, as previously, \(c_k\) is generated by the encryption transformation \(e\) which combines the running key \(K_k\) and the plaintext \(m_k\).

2.1.2 Receiver and reconstruction of the plaintext

For the usual stream cipher, the reconstruction of the plaintext requires the synchronization of the sequences of the running keys at the transmission and at the reception. At the receiver side, the decryption process is described, in the synchronous case, by:

\[
\begin{cases}
\hat{K}_k = f_{\theta}(\hat{K}_{k-1}) \\
\hat{m}_k = d(\hat{K}_k, c_k)
\end{cases}
\]

(3)

and, in the self-synchronous case, by:

\[
\begin{cases}
\hat{K}_k = f_{\theta}(c_{k-1}, \ldots, c_{k-t}) \\
\hat{m}_k = d(\hat{K}_k, c_k)
\end{cases}
\]

(4)

In both cases, the decryption transformation \(d\) is such that \(\hat{m}_k = m_k\) if \(\hat{K}_k = K_k\). For the synchronous stream cipher, the sequence \(\{K_k\}\) resulting from autonomous recurrences, the key generators \(f_{\theta}\) at both sides have to be initialized at the same value \(K_0 = K_0\). This initial value \(K_0\) can be considered as a static key. At the contrary, for the self-synchronous stream cipher, the sequences of the running key synchronize automatically.

2.2 Message-Embedded Cryptosystem

2.2.1 Transmitter and encryption

Generally, the message-embedded cryptosystem [3] obeys, at the transmitter side:
\[ \sum_\Theta \left\{ \begin{array}{l}
  x_{k+1} = f_\Theta(x_k, m_k) \\
  y_k = h_\Theta(x_k, m_k)
\end{array} \right. \]

where \( x_k \in \mathbb{R}^n, m_k \in \mathbb{R} \) and \( y_k \in \mathbb{R} \). \([m_k]\) means that \( h_\Theta \) can depend on \( m_k \) but not necessary. The principle of the message embedded cryptosystem is illustrated on Figure 2.

![Figure 2: Message-Embedded Cryptosystem](image)

Each symbol \( m_k \) is embedded in a sequence \( \{x_k\} \) generated by a nonlinear chaotic map \( f_\Theta \), where \( \Theta \in \Theta \) is a constant parameter. The most common nonlinearities \( f_\Theta \) are of polynomial type (Henon map, Logistic map, ...), or of linear piecewise type (Markov map). Only the quantity \( y_k \) is available at the transmitter output, \( x_k \) being an internal state which is not directly transmitted to the receiver.

### 2.2.2 Receiver and reconstruction of the plaintext

At the receiver side, the decryption process is described by:

\[ \left\{ \begin{array}{l}
  \hat{x}_{k+1} = g_{\hat{\Theta}}(\hat{x}_k, y_k) \\
  \hat{m}_k = d(\hat{x}_k, y_k)
\end{array} \right. \]

where \( y_k \) denotes a “window” of delayed outputs and of length to be determined.

The decryption transformation \( d \) is such that \( \hat{m}_k = m_k \) if \( \hat{x}_k = x_k \). However, it is not necessary here that the key generators at both sides are initialized to the same initial state \( x_0 \). Indeed, \( g \) is chosen such that, if \( \hat{\Theta} = \Theta \), then \( \hat{x}_k = x_k \), for all \( \hat{x}_0 \) and independently of \( m_k \). More precisely, we ensure either an asymptotic convergence:

\[ \lim_{k \to \infty} ||x_k - \hat{x}_k|| = 0 \quad \forall \hat{x}_0, \forall m_k \] (7)

or a finite-time convergence:

\[ \exists \delta_f, \|x_k - \hat{x}_k\| \leq \delta_f, \forall \hat{x}_0, \forall m_k, \forall k > k_f \] (8)

Particular structures of \( g \) ensuring a so-called Information Independent Global Synchronization (IIGS) have been introduced for encryption purposes in [3][9][10]. In [3][10], the synchronization of \( x_k \) and \( \hat{x}_k \) is formulated as a state reconstruction problem. An unknown input observer is proposed for \( g \) in (6). Its design is recalled below.

We consider a particular structure for the system (5), where \( f_\Theta \) and \( h_\Theta \) are some functions characterized by the matrices, \( \Theta_0 \in \mathbb{R}^{n \times n}, B_\Theta \in \mathbb{R}^{n \times 1} \) and \( C_\Theta \in \mathbb{R}^{1 \times n} \), such that:

\[ \left\{ \begin{array}{l}
  x_{k+1} = \Theta_0(p_k)x_k + B_\Theta m_k \\
  y_k = C_\Theta x_k
\end{array} \right. \] (9)

where \( p_k = q(y_k) \) with \( q \) a nonlinear function of \( y_k \). These systems are known as LPV (Linear Parameter Varying) systems and \( p_k \) is assumed to be available through the output. The matrices depend on a constant parameter vector \( \Theta = [\Theta_1, ..., \Theta_l]^T \in \Theta \). Actually, this structure is not very conservative because a specificity of usual chaotic systems including a nonlinearity of polynomial or linear piecewise type is that \( \Theta_0 \) can always be expressed in the following polytopic form [3]:

\[ \Theta_0(p_k) = \sum_{i=1}^{N} e_i(p_k)A_i \] (10)

where the \( A_i \)'s are constant matrices.

In order to ensure the reconstruction of the internal state \( x_k \), the function \( g \) has a structure of an unknown input polytopic observer, described by:

\[ \hat{x}_{k+1} = (P \Theta_0 - \mathcal{X} C_\Theta) \hat{x}_k + \mathcal{X} y_k + Q y_{k+1} \] (11)

where \( P, \mathcal{X} \) and \( Q \) are gain matrices which have to fulfill (7) or (8).

The stability conditions of the scheme (9)-(11) are given by the following theorem:

**Theorem 1** [3] The encryption scheme (9)-(11) is IIGS whenever the following conditions are satisfied:

- \( \text{rank}(CB) = \text{rank}(B) = 1 \)
there exists $S_1 > 0$, $S_j > 0$, $G_i$ and $F_i$, such that the following set of linear matrix inequalities:

$$
\begin{bmatrix}
G_i + G_i^T - S_i & G_i^T PA_{\theta}^{(i)} - F_i^T C_{\theta} \\
( PA_{\theta}^{(i)} )^T G_i - C_{\theta}^T F_i & S_j
\end{bmatrix} > 0
$$

is feasible $\forall (i, j) \in \{1, \ldots, N\} \times \{1, \ldots, N\}$.

**Remark 1** The rank condition means that the relative degree of (9) has to be equal to 1, which is not conservative in practice.

In [11], it has been proved that the conservatism of the stability conditions (12) can be reduced by considering the vertices $A_{\theta}^{(i)}$ of the minimal polytope including $p_k$, which are functions of the chaotic attractor. This, the polytopic observer can be an alternative, for example, to the well-known Extended Kalman Filter which does not take into account the chaos specificities.

The reconstruction of the information $m_k$ at each iteration $k$ is then given by:

$$
\hat{m}_k = (C_{\theta} B_{\theta})^{-1} (y_{k+1} - C_{\theta} x_{\theta} \delta_k)
$$

and $\hat{m}_k = m_k$ if $\delta_k = x_k$.

According to (13), the reconstruction of the plaintext $m_k$ requires the knowledge of the internal state vector $x_k$ although it is not transmitted. Since $x_k$ is essential to retrieve $m_k$, $x_k$ plays the role of a running key. To reconstruct $x_k$ from (11), it is clear that the system parameter $\theta$ is necessary and $\theta$ is considered as the static key.

An advantage of this method compared to the usual stream cipher is that whenever the synchronization is lost accidentally, an automatic resynchronization is ensured.

## 3 Cryptanalysis

Cryptanalysis is the science of studying attacks against cryptosystems in order to reveal their possible weakness.

For the standard stream cipher, according to (3) for the synchronous case and to (4) for the self-synchronous case, the plaintext $m_k$ reconstruction requires the synchronization of the sequences $\{K_k\}$ and $\{\hat{K}_k\}$ acting as the running keys.

The generation of these sequences by LFSR is a common mechanism for synchronous stream cipher. Nevertheless, Massey [12] has proved that the reconstruction of a whole sequence can be achieved from the knowledge of a fragment whose length is fixed and linked to the linear complexity of the LFSR. This is an example of weakness. Indeed, the pseudo-random sequence can be retrieved by carrying out a known plaintext attack, which consists in choosing a segment of the plaintext $m_k$ and in analysing the corresponding ciphertext $c_k$. Thus, it is worth noting that forcing $m_k$ to $0$ in (1) or in (2), $e$ being the XOR operation, we obtain $\hat{K}_k = c_k$ which is available at the output. The analysis of the sequence $\{K_k\}$ can then allow to retrieve the initial condition $K_0$ of $\hat{K}_k$ (see 2.1.2), which is the static key.

For the message-embedded cryptosystem, we consider the so-called known plaintext attack, i.e., the eavesdropper is supposed to know a sequence of $\{m_k\}$ and the corresponding $\{y_k\}$. To retrieve $m_k$, an eavesdropper cannot analyse the sequence $\{x_k\}$ since, even if $m_k = 0$, $x_k$ does not appear at the output. We admit that the eavesdropper has no other strategy than trying all the possible static keys $\theta$ (brute force attack or exhaustive search) and analyzing the corresponding pairs $(m_k, y_k)$. Thus, the most difficult situation for the eavesdropper is that there exists a unique $\theta$ that generates $y_k$ from $m_k$. The unicity of $\theta$ can be formulated in terms of the parametric identifiability whose some basics are recalled below.

The following definition is borrowed from [13].

**Definition 1** The system $\Sigma_\theta$ is structurally globally identifiable if for almost any $\hat{\theta} \in \theta$, $\Sigma_{\hat{\theta}} = \Sigma_\theta \Rightarrow \hat{\theta} = \theta$.

The system $\Sigma_\theta$ is structurally locally identifiable if for almost any $\hat{\theta} \in \theta$, there exists a neighbourhood $v(\theta)$ such that $\hat{\theta} \in v(\theta)$ and $\Sigma_{\hat{\theta}} = \Sigma_\theta \Rightarrow \hat{\theta} = \theta$.

The system $\Sigma_\theta$ is structurally unidentifiable if for almost any $\hat{\theta} \in \theta$, there is no neighbourhood $v(\theta)$ such that $\hat{\theta} \in v(\theta)$ and $\Sigma_{\hat{\theta}} = \Sigma_\theta \Rightarrow \hat{\theta} = \theta$.

Several methods for testing the parametric identifiability exist as the Taylor series expansion [14], the local state isomorphism approach [15] and the Gröbner bases
method [16]. We only describe here the local state isomorphism approach and the Gröbner bases method.

3.1 Local state isomorphism approach

Consider the general system expressed as:

\[
\begin{align*}
x_{k+1} &= f_\theta^{(1)}(x_k) + m_k f_\theta^{(2)}(x_k) \\
y_k &= h(x_k)
\end{align*}
\]  

(14)

where \( f_\theta^{(1)} \) and \( f_\theta^{(2)} \) are nonlinear functions of \( x_k \) and \( \theta = [\theta_1, \ldots, \theta_L]^T \in \Theta. \) The initial condition of \( x_k \) is denoted by \( x_0(\theta). \)

**Remark 2** For the systems of the form (9) considered in the Section 2.2.2, \( f_\theta^{(2)} = B_\theta \) where \( B_\theta \) is a constant matrix.

Consider the system (14). Assume that \( \Sigma_\theta \) is locally reduced at \( x_0(\theta) \) for almost any \( \theta \in \Theta, \) that is, it satisfies both the controllability rank condition and the observability rank condition [17]. The following proposition establishes a condition for global identifiability of system (14), as a discrete counterpart of the theorem found in [15].

**Proposition 1** \( \Sigma_\theta \) and \( \Sigma_{\hat{\theta}} \) have the same input-output behavior for any \( m_k \) if and only if there exists a local state isomorphism \( \phi, \) defined by \( x_k \in \mathbb{R}^n \to \phi(x_k) \in \mathbb{R}^n, \) such that, for any \( x_k \) in the neighborhood \( \nu(x_0), \) the following conditions are satisfied:

\[
\begin{align*}
(i) \quad \text{rank}(\frac{\partial \phi(x_k)}{\partial x_k}) &= n, \\
(ii) \quad \phi(x_0(\hat{\theta})) &= x_0(\theta), \\
(iii) \quad f_\theta^{(1)}(\phi(x_k)) &= \frac{\partial \phi(x_k)}{\partial x_k} f_\theta^{(1)}(x_k), \\
(iv) \quad f_\hat{\theta}^{(2)}(\phi(x_k)) &= \frac{\partial \phi(x_k)}{\partial x_k} f_\theta^{(2)}(x_k), \\
(v) \quad h_\hat{\theta}(\phi(x_k)) &= h_\theta(x_k).
\end{align*}
\]

(15)

These conditions express that \( \phi \) is a diffeomorphism (i), the initial states coincide (ii), the dynamic terms coincide (iii), the control terms coincide (iv) and the observation terms coincide (v). After checking that \( \Sigma_\theta \) is locally reduced at \( x_0(\theta), \) one can look for all solutions for \( \hat{\theta} \) and \( \phi \) of (15). If, for almost any \( \theta, \) the only possible solution is \( \hat{\theta} = \theta \) and \( \phi(x_k) = x_k, \) then \( \Sigma_\theta \) is globally identifiable.

In the case \( f_\theta^{(1)} \) and \( f_\theta^{(2)} \) are polynomials in \( x_k, \) parameterized by \( \theta, \) and \( h_\theta(x_k) = C_\theta x_k, \) \( \phi \) can directly be written as a linear transformation \( \phi(x_k) = T x_k, \) which simplifies the calculations. Hence, the Proposition 1 turns into:

**Proposition 2** \( \Sigma_\theta \) and \( \Sigma_{\hat{\theta}} \) have the same input-output behavior for any \( m_k \) if and only if there exists a linear transformation \( T \) such that the following conditions are satisfied:

\[
\begin{align*}
(i) \quad \text{det}(T) \neq 0, \\
(ii) \quad T x_0(\hat{\theta}) &= x_0(\theta), \\
(iii) \quad f_\theta^{(1)}(T x_k) &= T f_\theta^{(1)}(x_k), \\
(iv) \quad f_\hat{\theta}^{(2)}(T x_k) &= T f_\theta^{(2)}(x_k), \\
(v) \quad C_\theta T &= C_\theta.
\end{align*}
\]

If, for almost any \( \theta, \) the only possible solution is \( \hat{\theta} = \theta \) and \( T = I \) where \( I \) is the identity matrix of dimension \( n, \) then \( \Sigma_\theta \) is globally identifiable.

Note that Propositions 1 and 2 are just conjectures.

Another approach to test the parametric identifiability is the Gröbner bases method, exposed in the next section.

3.2 Gröbner bases approach

In order to test the parameter identifiability, we want to obtain an input/output relation, with the general form:

\[
L_\theta(y_k, y_{k+1}, \ldots, m_k, m_{k+1}, \ldots) = 0
\]

(17)

with \( L_\theta, \) a function parameterized by \( \theta. \) To this end, the internal state \( x_k \) must be eliminated and
hence is considered as indeterminate. The elimination can be achieved thanks to the method of the Gröbner bases, borrowed from algebra. The first algorithm of this type is due to [16]. Some notions of differential algebra can be found in [18] for continuous-time systems. However, they can equally be defined with the derivative operator (continuous-time case) or with the delay operator (discrete-time case). Some recalls in the case of discrete-time systems are carried out below.

Consider a system $\Sigma_0$ of the polynomial ring, denoted by $\mathcal{A} = \mathbb{R}[x_k^{(1)}, \ldots, x_k^{(n)}]$ where the indeterminates are $x_k^{(1)}, \ldots, x_k^{(n)}$ and the coefficients are real numbers.

**Definition 2** An ideal of $\mathcal{A}$ is a subset $I$ of $\mathcal{A}$, such that:

- $\forall p \in I, \forall q \in I, p + q \in I$
- $\forall p \in I, \forall g \in I, pg \in \mathcal{A}$

**Definition 3** The ideal, generated by the system with polynomial nonlinearities $\Sigma_0$ (5), in $\mathcal{A}$, is the set of all linear combinations of the elements of $\Sigma_0$ with any elements of $\mathcal{A}$ for coefficients.

**Definition 4** A lexicographic order is a ranking according to the names of the variables and their iterates such that:

- $x_k^{(i)} < x_{k+l}^{(i)}, \forall l \in \mathbb{N}^+$,
- $x_k^{(i)} < x_{k+i}^{(i)} \Rightarrow x_k^{(i)} < x_{k+i}^{(i)}, \forall l \in \mathbb{N}^+, \forall l \in \mathbb{N}^+$,
- $x_k^{(i)} < x_k^{(j)} \Rightarrow (x_k^{(i)})^\alpha < (x_k^{(j)})^\beta, \forall \alpha \in \mathbb{N}^+, \forall \beta \in \mathbb{N}^+$

The variables to be eliminated are considered as the greatest.

If a given pair $(m_k, y_k)$ satisfies (5), it will also satisfy equations that are obtained by addition and by multiplication of (5), that is the ideal generated by (5). For a given lexicographic order, it then suffices to find a basis of this ideal whose expressions do no longer contain the variables $x_k$, but only contain $y_k$, $m_k$, their iterates and $\theta$. These expressions of the basis are of the required form (17). Such a basis is called a Gröbner basis. A more formal definition of the Gröbner bases can be found in [16] and a theorem of variable elimination based on this method is detailed in [19]. After obtaining the relation (17) thanks to the Gröbner basis method, the following theorem formulates a necessary and sufficient condition for parameter global identifiability.

**Theorem 2** [20] The parameter vector $\theta$ is globally identifiable if and only if the equations (17) can be re-arranged in a linear regression such that, parameter by parameter:

$$P_i(y_k, m_k)\theta_i - Q_i(y_k, m_k) = 0 \quad i = 1, \ldots, L$$

where $P_i$ and $Q_i$ are polynomials depending only on $y_k$, $m_k$, and on their iterates, and $L = \dim(\theta)$.

Discussion

3.2.1 Brute force attack

If the equation (17) admits several possible solutions for the parameter $\theta$, $\theta$ is not identifiable. In this case, an eavesdropper has a favorable chance to find $\theta$ by a brute force attack because several solutions are possible for $\theta$. Thus, the parameter $\theta$ is a bad candidate to play the role of the static key.

If the equation (17) admits a unique solution for $\theta$, then it is more difficult for the eavesdropper to find $\theta$ from (17) by an exhaustive search. Consequently, the parameter $\theta$ may be a good candidate to play the role of the static key against a brute force attack.

3.2.2 Algebraic attack

Contrarily to the brute force attack, if the eavesdropper knows the structure of the algorithm, (20) highlights the fact that he is able to retrieve easily the parameters. Indeed, he must solve a system with $L$ linear equations with $L$ unknowns. Solving (20) is a kind of algebraic attack.

A fundamental conclusion derived from this analysis is that cryptosystems involving only polynomial nonlinearities (for which Gröbner bases are dedicated) are weak against algebraic attacks.
4 Illustrative examples

4.1 Example 1

Consider the message-embedded cryptosystem which obeys, at the transmitter side:

\[
\begin{align*}
    &x^{(1)}_{k+1} = ax^{(1)}_k - bx^{(2)}_k \\
    &x^{(2)}_{k+1} = bx^{(1)}_k + \theta_1 x^{(2)}_k + \theta_2 (x^{(2)}_k)^2 + \theta_3 (x^{(2)}_k)^3 + m_k \\
    &y_k = x^{(2)}_k
\end{align*}
\]

where \( m_k \) represents the plaintext.

4.1.1 Local state isomorphism approach

It can be shown that system (21) is locally reduced at \( x_0(\theta) \) for all \( \theta \in \Theta \). The system (21) has polynomial nonlinearities and \( h_0(x_k) = x^{(2)}_k \). Hence, the local state isomorphism \( \phi \) can be written as a linear transformation \( \phi(x_k) = Tx_k \). Let define the matrix \( T \), with \( t_i \in \mathbb{R} \), as:

\[
T = \begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix}
\]

with \( \det(T) \neq 0 \). The conditions (16)-(v) and (16)-(iv) implies respectively that \( t_3 = 0, t_4 = 1 \) and, \( t_2 = 0 \). Condition (16)-(iii) implies:

\[
a t_1 x^{(1)}_k - b x^{(2)}_k = t_1 (a x^{(1)}_k - b x^{(2)}_k)
\]

\[
b x^{(1)}_k + \Delta_1 x^{(2)}_k + \Delta_2 (x^{(2)}_k)^2 + \Delta_3 (x^{(2)}_k)^3 =
\]

\[
= b x^{(1)}_k + \theta_1 x^{(2)}_k + \theta_2 (x^{(2)}_k)^2 + \theta_3 (x^{(2)}_k)^3
\]

The equation (23) leads to \( t_1 = 1 \). The equation (24) is equivalent to:

\[
(\hat{\theta}_1 - \theta_1) x^{(2)}_k + (\hat{\theta}_2 - \theta_2) (x^{(2)}_k)^2 + (\hat{\theta}_3 - \theta_3) (x^{(2)}_k)^3 = 0
\]

which implies that \( \hat{\theta}_1 = \theta_1, \hat{\theta}_2 = \theta_2 \) and \( \hat{\theta}_3 = \theta_3 \), assuming that \( x^{(2)}_k \neq 0 \). Consequently, the matrix \( T \) reduces to the identity matrix of dimension 2. Hence, \( \hat{\theta}_1, \hat{\theta}_2 \) and \( \hat{\theta}_3 \) are globally identifiable.

4.1.2 Gröbner bases approach

Since \( x^{(1)}_k \) is not directly transmitted through the signal \( y_k \), it is chosen to be the greatest and the corresponding lexicographic order is:

\[
x^{(2)}_k < x^{(2)}_{k+1} < x^{(2)}_k < x^{(1)}_k < x^{(1)}_{k+1}
\]

The ideal of the Gröbner basis generated by the system (21), with the lexicographic order (26), is:

\[
y_k + y_{k+1} + \theta_1 (ay_k - y_{k+1}) + \theta_2 (ay_k^2 - y_{k+1}^2) + \theta_3 (ay_k^3 - y_{k+1}^3) - ay_{k+1} - b^2 y_k + am_k - m_{k+1} = 0
\]

By iterating the equation (27), we get a system of linear equations with three unknowns \( \hat{\theta}_1, \hat{\theta}_2 \) and \( \hat{\theta}_3 \). We can then write three expressions of the form:

\[
P_1(y_k, m_k) \hat{\theta}_1 - Q_1(y_k, m_k) = 0
\]

\[
P_2(y_k, m_k) \hat{\theta}_2 - Q_2(y_k, m_k) = 0
\]

\[
P_3(y_k, m_k) \hat{\theta}_3 - Q_3(y_k, m_k) = 0
\]

So, Theorem 2 is fulfilled for each parameter and the same conclusion is reached as with the local state isomorphism approach: the parameters \( \hat{\theta}_1, \hat{\theta}_2 \) and \( \hat{\theta}_3 \) are globally identifiable. Hence, against brute force attack, \( \hat{\theta}_1, \hat{\theta}_2 \) and \( \hat{\theta}_3 \) may play the role of the static key. On the other hand, it is easy to solve the system (28), highlighting the weakness of the cryptosystem (21) against algebraic attack.

4.2 Example 2

Consider now the message-embedded cryptosystem where the plaintext \( m_k \) is embedding in the Henon map:

\[
\begin{align*}
    &x^{(1)}_{k+1} = \theta_1 (x^{(1)}_k) + \theta_2 x^{(2)}_k + m_k \\
    &x^{(2)}_{k+1} = \theta_3 x^{(1)}_k + \theta_4 m_k \\
    &y_k = x^{(1)}_k
\end{align*}
\]

This example is only treated through the Gröbner bases method.

Since \( x^{(2)}_k \) is not directly transmitted through the signal
$y_k$, it is chosen to be the greatest and the corresponding lexicographic order is:

\[ x^{(1)}_k < x^{(1)}_{k+1} < x^{(2)}_k < x^{(2)}_{k+1} < x^{(2)}_{k+2} \]  

(30)

The ideal of the Gröbner basis generated by the system (29), with the lexicographic order (30), is:

\[
\theta_1 y_{k+1}^2 + \theta_2 \theta_3 y_k - y_{k+2} + m_{k+1} + \theta_2 \theta_4 m_k = 0 \tag{31}
\]

By iterating the equation (31), we get three expressions of the form:

\[
P_1(y_k, m_k) \theta_1 - Q_1(y_k, m_k) = 0
\]

\[
P_2(y_k, m_k) \theta_2 \theta_3 - Q_2(y_k, m_k) = 0
\]

\[
P_3(y_k, m_k) \theta_2 \theta_4 - Q_3(y_k, m_k) = 0
\]

(32)

Theorem 2 is satisfied for $\theta_1$ and the products $\theta_2 \theta_3$ and $\theta_2 \theta_4$, $\theta_1$, the products $\theta_2 \theta_3$ and $\theta_2 \theta_4$ are globally identifiable, but not the parameters $\theta_2$, $\theta_3$ and $\theta_4$ themselves. As there is several possible solutions for $\theta_2$, $\theta_3$ and $\theta_4$, they are bad candidates for the static key against brute force attacks.

5 Conclusion

This paper has carried out a comparison between two encryption schemes, the standard stream cipher and the message-embedded cryptosystem. The comparison was focused, on one hand, on the synchronization of the running key sequences and, on the other hand, on the cryptanalysis of the encryption algorithms. The cryptanalysis has been studied through the reconstruction of the transmitter static parameter. A formalism based on the parametric identiability has been proposed in the case of the message-embedded cryptosystem. As a result, the identifiable parameters are good candidates to play the role of the static key against brute force attack. However, they can easily be retrieved by performing an algebraic attack. A fundamental conclusion is that the cryptosystems involving only polynomial nonlinearities are weak against algebraic attacks.

References


