Cryptanalysis of a family of self-synchronizing chaotic stream ciphers

David Arroyo*[,a], Gonzalo Alvarez[a], José María Amigó[b], Shujun Li[c]

[a] *Instituto de Física Aplicada, Consejo Superior de Investigaciones Científicas, Serrano 144, 28006 Madrid, Spain*
[b] *Centro de Investigación Operativa, Universidad Miguel Hernández, Avda. de la Universidad s/n, 03202 Elche, Spain*
[c] *Fachbereich Informatik und Informationswissenschaft, Universität Konstanz, Fach M697, Universitätsstraße 10, 78457 Konstanz, Germany*

Abstract

Unimodal maps have been broadly used as a base of new encryption strategies. Recently, a stream cipher has been proposed in the literature, whose keystream is basically a symbolic sequence of the (one-parameter) logistic map or of the tent map. In the present work a thorough analysis of the keystream is made which reveals the existence of some serious security problems.

Key words: Unimodal maps, symbolic dynamics, stream cipher, known-plaintext attack, control parameter estimation, initial condition estimation.

1. Introduction

The partition of the state space transforms a measure-preserving dynamical system into a stationary stochastic process called a symbolic dynamics. In the case of chaotic systems (i.e., governed by chaotic maps), the resulting symbolic dynamics has some specific properties, like sensitivity to initial conditions and strong mixing, which are very attractive for cryptographic purposes. For instance, the symbolic dynamics of a chaotic map can be used as a Random Number Generator (RNG) [1] and, more generally, as a source of entropy. Unimodal maps are particularly useful in this regard, since their generating partitions comprise two intervals, thus leading to a natural source of Random Bit Generators (RBGs). Among all possible applications of RNGs and RBGs in cryptography, their role as keystream generators in stream ciphers is especially important.

Recently a stream cipher based on the symbolic dynamics of the (one-parameter) logistic map and tent map, was proposed in [2]. If the parameter of either map is selected conveniently, then its symbolic sequences pass all the statistical tests necessary for their consideration as keystreams. However, we show that this requirement is not enough to guarantee the security of this stream cipher and point out some cryptographic weaknesses.

The work described in this paper is organized as follows. In Sec. 2 the encryption scheme of [2] is explained. After that, some issues relevant to the security of the cryptosystem are highlighted (Sec. 3). In Sec. 4 the cryptosystem is analyzed taking into account the dynamics of the underlying chaotic system. The problems derived from the selection of the logistic and tent maps are also discussed there. Finally, the main results and conclusions of the work are summarized in Sec. 5.

2. Description of the encryption scheme

In the cryptosystem described in [2] the transformation of the plaintext into the ciphertext is done bitwise and driven by symbolic sequences generated either by the logistic map or by the tent map. Recall that the *logistic map* is defined as

$$f_\lambda(x) = \lambda x(1-x), \tag{1}$$

*Corresponding author: David Arroyo (david.arroyo@iec.csic.es).

for $x \in [0,1]$ and $\lambda \in [0,4]$, and the *tent map* is given by the following equation:

$$f_\lambda(x) = \begin{cases} x/\lambda, & \text{if } 0 \le x < \lambda, \\ (1-x)/(1-\lambda), & \text{if } \lambda \le x \le 1, \end{cases} \tag{2}$$

where $x \in [0,1]$ and $\lambda \in (0,1)$. Henceforth we will refer to λ as the control parameter.

Given a closed interval $I \subset \mathbb{R}$ and a map $f : I \to I$, the *orbit* of (the initial condition) $x \in I$ is defined as the set $\mathcal{O}_f(x) = \{f^n(x) : n \in \mathbb{N}_0\}$, where $\mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0, 1, ...\}$, $f^0(x) = x$ and $f^n(x) = f\left(f^{n-1}(x)\right)$. A continuous map defined on an interval of $\mathbb{R}$ that is increasing (decreasing) to the left of an interior point and decreasing (increasing) thereafter is said to be *unimodal*. Examples of unimodal maps are provided by the logistic and tent maps. A unimodal map attains its maximum (minimum) at a single point x_c ($x_c = 1/2$ for the logistic map, and $x_c = \lambda$ for the tent map), called the *critical point*. If $f : I \to I$ is unimodal, then any orbit can be encoded into a binary sequence,

$$\mathbf{B}_\infty(f, x) = \{B_i(f, x)\}_{i=0}^{\infty} = \theta(f^0(x))\theta(f^1(x)) \dots \theta(f^{N-1}(x))..., \tag{3}$$

where $\theta(\cdot)$ is the step function

$$\theta(y) = \begin{cases} 0, & \text{if } y < x_c, \\ 1, & \text{if } y \ge x_c. \end{cases} \tag{4}$$

In [2] the plaintext is encrypted through the symbolic dynamics of either the logistic map or the tent map, with fixed control parameter λ and initial condition x_0. If the plaintext is N bit long, then the first $m + N$ points of $\mathcal{O}_{f_\lambda}(x_0)$ are computed with the selected map, and the corresponding symbolic sequence is produced. The first m bits of this symbolic sequence are used to bear the initial condition. Indeed, according to the theory of symbolic dynamics, given $\varepsilon > 0$ and a generating partition α of I with respect to f (like the partition $\alpha = \{[0, x_c), [x_c, 1]\}$ in the case of unimodal maps of the unit interval), any real number $x \in I$ can be represented with precision ε as a symbolic sequence of f with respect to α, with initial condition x and length above a certain threshold [3]. Therefore, once the precision ε has been set, m is chosen to be larger than the corresponding threshold. The scheme proposed in [2] divides the finite binary sequence

$$\mathbf{B}_{m+N}(f_\lambda, x_0) = \{B_i(f_\lambda, x_0)\}_{k=0}^{m+N-1} = \theta(f^0(x))\theta(f^1(x)) \dots \theta(f^{m+N-1}(x))$$

into two segments: $\mathbf{B}^{\text{init}} = \{B_i^{\text{init}}\}_{i=0}^{m-1}$ with $B_i^{\text{init}} = B_i(f_\lambda, x_0)$, and $\mathbf{B}^{\text{ks}} = \{B_i^{\text{ks}}\}_{i=0}^{N-1}$ with $B_i^{\text{ks}} = B_{m+i}(f_\lambda, x_0)$. The initial segment $\mathbf{B}^{\text{init}}$ contains the information on x_0 up to the precision wished. The final segment $\mathbf{B}^{\text{ks}}$ is the keystream of the cipher, i.e., the plaintext $\mathbf{P} = \{P_i\}_{i=0}^{N-1}$ is transformed into the *pre-ciphertext* $\mathbf{C} = \{C_i\}_{i=0}^{N-1}$ according to

$$C_i = P_i \oplus B_i^{\text{ks}} = P_i \oplus B_{m+i}(f_\lambda, x_0), \tag{5}$$

where $i = 0, 1, \dots, N-1$, and $0 \oplus 0 = 1 \oplus 1 = 0$, $0 \oplus 1 = 1 \oplus 0 = 1$. Finally, the pre-ciphertext $\mathbf{C}$ and $\mathbf{B}^{\text{init}}$ are combined into the *ciphertext* or encrypted message $\mathbf{M} = \{M_i\}_{i=0}^{m+N-1}$ which is sent to the receiver through an insecure channel. The generation of $\mathbf{M}$ is driven by a shuffler block, implementing an injective map $\pi : \{0, 1, \cdots, m-1\} \mapsto \{0, 1, \cdots, N-1, N\}$, which inserts the m bits of $\mathbf{B}^{\text{init}}$ into the pre-ciphertext $\mathbf{C}$ according to the following rule. (a) If $0 \le \pi(i) \le N-1$, then B_i^{init} is inserted before $C_{\pi(i)}$; (b) if $\pi(i) = N$, then B_i^{init} is inserted after C_{N-1}, i.e., $M_{m+N-1} = B_i^{\text{init}}$. Thus, a ciphertext with, say, $\pi(i) < N$ for all i, looks as follows:

$$\mathbf{M} = C_0 C_1 \cdots C_{i_0-1} B_{j_0}^{\text{init}} C_{i_0} \cdots C_{i_{m-1}-1} B_{j_{m-1}}^{\text{init}} C_{i_{m-1}} \cdots C_{N-1}, \tag{6}$$

where $\pi(j_k) = i_k$, $k \in \{0, 1, \cdots, m-1\}$, and $i_0 < ... < i_{m-1}$. The shuffler block, i.e., the map π, is also known at the receiver, thus making the recovery of x_0 feasible.

In sum, the encryption is done in three steps:

(1) *Symbolic sequence*: $\mathbf{B}_{m+N}(f_\lambda, x_0) = \mathbf{B}^{\text{init}} \big\| \mathbf{B}^{\text{ks}}$ ($\|$ stands for "juxtaposition").

(2) *Pre-ciphertext*: $\mathbf{C} = \mathbf{P} \oplus \mathbf{B}^{\text{ks}}$ (the $\oplus$ operation is bitwise)

(3) *Ciphertext*: $\mathbf{M} = \pi\left(\mathbf{B}^{\text{init}} \| \mathbf{C}\right)$ (abusing notation, $\pi(\mathbf{S}_1\|\mathbf{S}_2)$ stands here and henceforth for the action of the shuffling map π on the binary sequence $\mathbf{S}_1\|\mathbf{S}_2$ of length $m+N$, as exemplified in Eq. (6)).

In order to decrypt $\mathbf{M}$, the receiver extracts $\mathbf{B}^{\text{init}}$ from $\mathbf{M}$ to determine x_0; the remaining bits form $\mathbf{C}$. This allows the receiver to replicate $\mathbf{B}^{\text{ks}}$ by computing the orbit of x_0 under the selected chaotic map, using the right value of the control parameter. Lastly, the plaintext is recovered as

$$P_i = C_i \oplus B_i^{\text{ks}} = C_i \oplus B_{m+i}(f_\lambda, x_0), \tag{7}$$

for $k = 0, 1, \ldots, N-1$.

An explicit definition of the key of the cryptosystem is not given in [2]. Nevertheless, in [2, Sec. 3] it is pointed out that the map selected (either the logistic map or the tent map), its control parameter, and the position of the bits of $\mathbf{B}^{\text{init}}$ in $\mathbf{M}$ are necessary to recover the plaintext. Henceforth, it is assumed that the key consists of these three elements or "subkeys". The map π used in the shuffler block might be given by an m-dimensional vector $\boldsymbol{\pi} = [\pi(0), \cdots, \pi(m-1)]$, where $\pi(i)$ is a $\lceil \log_2(N+1) \rceil$-bit integer. In practice, a secret seed s could be used to generate the map π in a pseudo-random manner; in this case, the subkey corresponding to the shuffler reduces to the seed s.

3. Design problems

3.1. Key space

The complete definition of a cryptosystem demands the precise and thorough specification of the set of values of the secret key [4, Rule 5]. As mentioned above, the control parameter and the initial condition of the chaotic map (necessary to build the pre-cipher text $\mathbf{C}$) are certainly part of the key. In relation with the control parameter, the considered maps must be evaluated to guarantee that they evolve chaotically during the encryption stage. In the case of the logistic map, the selection of adequate values for λ is quite complex since the bifurcation diagram of this map possesses a dense set of periodic windows [5]. Therefore, if the keystream has to be generated with the logistic map, one must assure that the Lyapunov exponent of f_λ is positive.

On the other hand, the tent map is not a good source for generating pseudo-random bits from its symbolic dynamics. Since the Lebesgue measure on $[0,1]$ is an ergodic invariant measure of the tent map for all $\lambda \in (0,1)$ [6], it follows that the ratio between the number of 1-bits and 0-bits in a typical orbit coincides with the ratio between the lengths of the intervals $[\lambda, 1]$ and $[0, \lambda)$, namely, $\frac{1-\lambda}{\lambda}$. Therefore, in order to have an approximately balanced bit sequence $\mathbf{B}^{\text{ks}}$, λ should be close to $1/2$.

3.2. Considerations about the synchronization procedure

In chaos-based cryptosystems, decryption of the ciphertext requires perfect regeneration of the orbit(s) involved in the encryption stage. This being the case, the receiver must know the control parameter(s) and the initial condition(s) used by the transmitter. Those values can be obtained by the receiver from either the secret key or the design specifications. However, the agreement on the initial condition can be settled *indirectly* using *synchronization techniques*. Indeed, if the chaotic systems at the transmitter and receiver are suitably coupled, their orbits converge to each other although they have been derived from different initial conditions. Synchronization implies that, after a transient time, the chaotic system(s) at the receiver reproduces the dynamics of the chaotic systems(s) at the transmitter, which further allows the recovering of the plaintext without knowledge of x_0. This is certainly not the case of the cryptosystem proposed in [2], since the initial condition have to be known in order to reproduce the keystream $\mathbf{B}^{\text{ks}}$. As a consequence, the whole ciphertext must be received before decryption can start while, in conventional synchronization schemes, decryption is progressively achieved during the reception of the ciphertext. We conclude that, what is called *synchronization* in [2] is rather a method to codify and send the initial condition, than a usual synchronization technique.

Furthermore, nothing is mentioned in [2] about how x_0 is obtained from $\mathbf{B}^{\text{init}}$. We briefly address this issue here. According to the theory of symbolic dynamics, a symbolic sequence of length L partitions the

state interval I into 2^L subintervals $I_j^{(L)}$, $1 \leq j \leq 2^L$, that is, $I = I_1^{(L)} \cup I_2^{(L)} \cdots \cup I_{2^L}^{(L)}$, with $I_j^{(L)} \cap I_k^{(L)} = \varnothing$ for $j \neq k$. The binary sequences of length L obtained for each $x \in I_j^{(L)}$ are the same. Therefore, a given symbolic sequence of length L singles out the subinterval $I_j^{(L)}$ its initial condition x_0 belongs to, what provides an estimation of x_0. The estimation error depends on the width of $I_j^{(L)}$, which in turn depends on the map considered. In the case of the symmetric tent map (i.e., the tent map with $\lambda = 1/2$), all the subintervals $I_j^{(L)}$ have width equal to $1/(2^L)$ (see Fig. 1). Hence if the first n bits of $\mathbf{B}_L(f_{1/2}, x_0)$ locate x_0 in the subinterval $[k/2^n, (k+1)/2^n]$, then the bit B_n determines whether x_0 belongs to either the left ($B_n = 0$) or the right ($B_n = 1$) half of that subinterval. Nevertheless, this dichotomic search cannot be done in a general case. Indeed, the subintervals $I_j^{(L)}$ associated to the logistic map and the tent map with $\lambda \neq 1/2$ are not equal-width. However, in [7] it is shown that the symbolic sequences of unimodal maps can be assigned a linear order. This linear order preserves the order of the corresponding initial conditions in $\mathbb{R}$ and can be used to estimate x_0 through a binary search procedure [8].

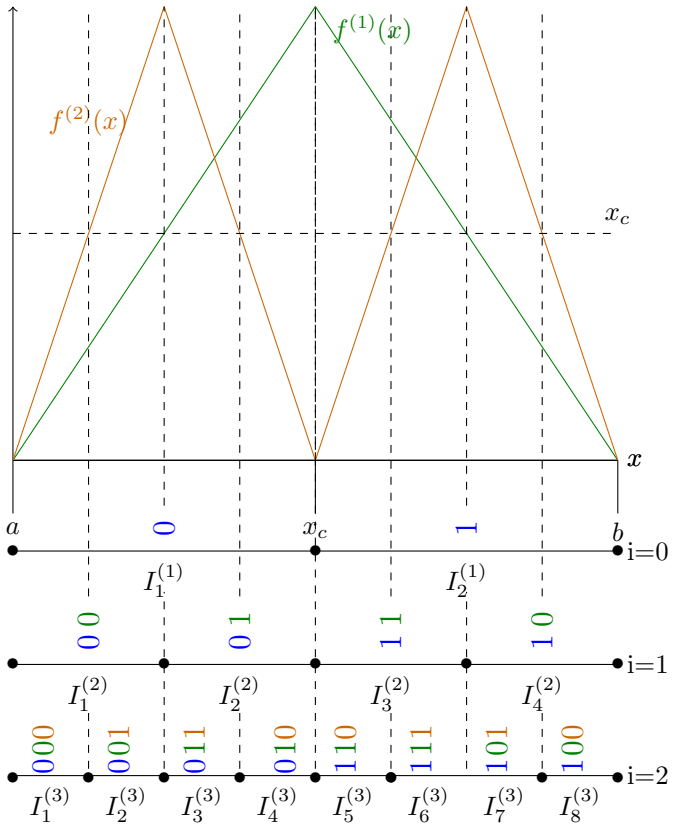


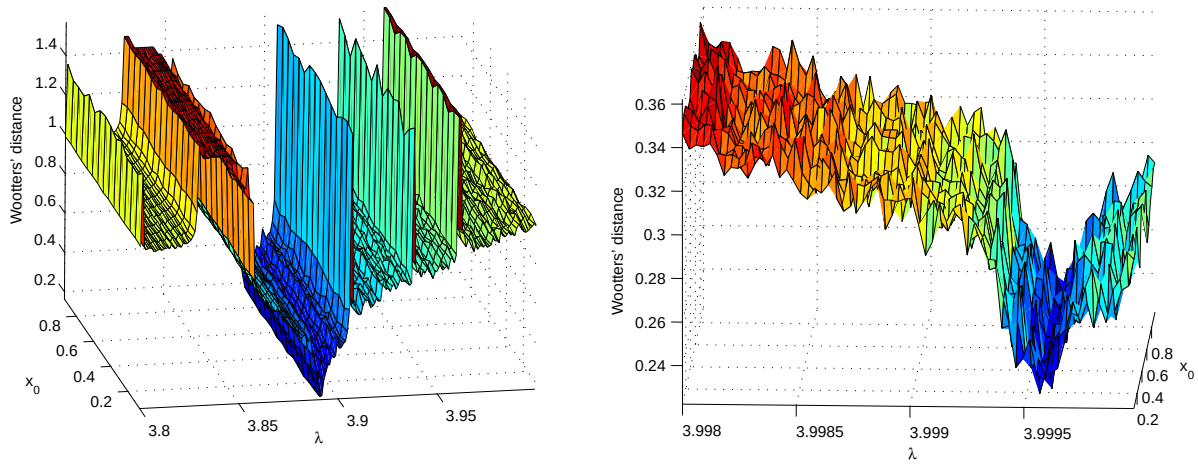
Figure 1: Symbolic intervals for different iterations of the symmetric tent map.

4. Problems derived from the dynamics of the underlying chaotic systems

A crucial step in the design of a chaos-based cryptosystem is the selection of the underlying chaotic map(s). In this section it is shown that the choice of the logistic map and the tent map for the scheme proposed in [2] implies serious security problems.

First of all, due to lack of details in [2], it is assumed that the interleaving of the symbolic block $\mathbf{B}^{\text{init}}$ in the pre-ciphertext $\mathbf{C}$ to build the ciphertext $\mathbf{M}$, is performed in a random way. In a *chosen-plaintext*

attack, a cryptanalyst has access to the encryption device and thus can obtain the output corresponding to any input. If $P_i = 0$ for $0 \leq i \leq N-1$, i.e., all the bits of the plaintext $\mathbf{P}$ are chosen to be zero, then $\mathbf{C} = \mathbf{0} \oplus \mathbf{B}^{\mathrm{ks}} = \mathbf{B}^{\mathrm{ks}}$, and the corresponding ciphertext is $\pi\left(\mathbf{B}^{\mathrm{init}} \| \mathbf{B}^{\mathrm{ks}}\right)$. Call this particular ciphertext $\mathbf{B}^{\mathrm{shuffled}}$. According to [9], given a symbolic sequence $\mathbf{B}_L(f_\lambda, x_0)$ of a unimodal map f_λ (see Eq.(3)), both the control parameter λ and the initial condition x_0 can be estimated in a straightforward way. Actually, the problem we are dealing with is not quite the same, since the available symbolic sequences are distorted through the permutation procedure. Nevertheless, we will presently show that the estimation of the control parameter is still possible using $\mathbf{B}^{\mathrm{shuffled}} = \pi\left(\mathbf{B}^{\mathrm{init}} \| \mathbf{B}^{\mathrm{ks}}\right)$ instead of $B_{m+N}(f_\lambda, x_0) = \mathbf{B}^{\mathrm{init}} \| \mathbf{B}^{\mathrm{ks}}$, where f_λ is the logistic map or the tent map, and the estimation method depends on f_λ. Consequently the first step in the cryptanalysis of this cipher calls for discerning the chaotic map used in the generation of $\mathbf{B}^{\mathrm{shuffled}}$, the encryption of $\mathbf{P} = \mathbf{0}$. Once this step has been completed, the next step is to estimate the control parameter.



(a) $\hat{\lambda} = 3.890925$ (b) $\hat{\lambda} = 3.999567$

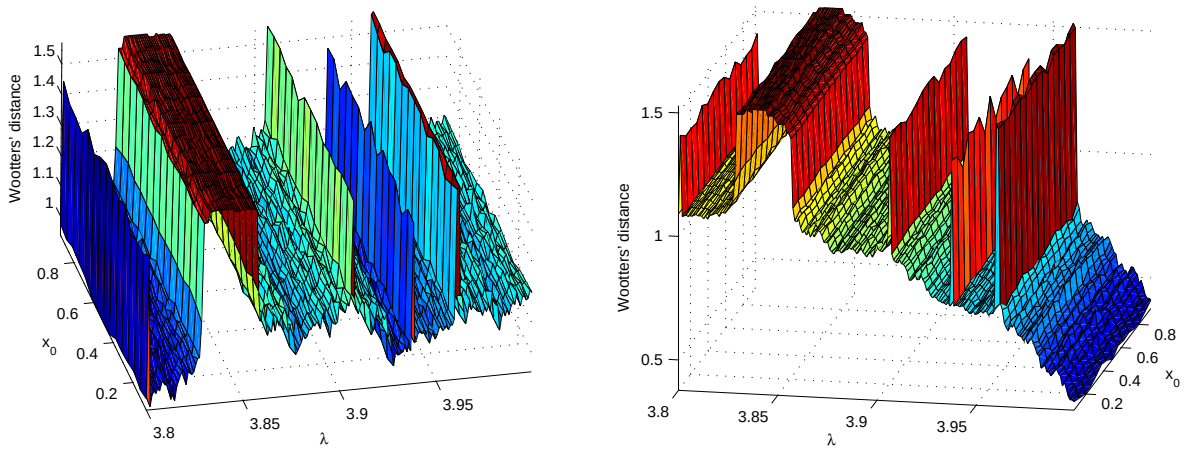
Figure 2: Wootters' distance of the logistic map with respect to the logistic map. The length of the symbolic sequences is $N = 10^4$, whereas the words are of width $w = 10$.

4.1. Identification of the chaotic map from symbolic sequences

The dynamics of every chaotic system has some particular characteristics that make it distinct. These "fingerprints" are also present in their symbolic dynamics and can be brought to light via statistical comparison of the corresponding symbolic sequences. A method along these lines exploits the "statistical distance" between symbolic sequences to discriminate one chaotic map from another. In this paper we consider the statistical distance defined by Wootters [10]. Let $P_i = \left\{p_j^{(i)}\right\}_{j=1}^{N}$ $(i = 1, 2)$ be two probability distributions. Wootters' statistical distance between P_1 and P_2 is given by

$$\mathcal{D}_W(P_1, P_2) = \cos^{-1}\left(\sum_{j=1}^{N} \sqrt{p_j^{(1)} \cdot p_j^{(2)}}\right). \tag{8}$$

Since $\mathcal{D}_W$ is calculated from two probability distributions, it is necessary to establish a method to derive a probability distribution from the dynamics of a unimodal map. Let $\mathbf{B}_L(f_\lambda, x_0)$ be a symbolic sequence of a unimodal map f_λ. A probability distribution can be obtained from $\mathbf{B}_L(f_\lambda, x_0)$ just by grouping all bits in a sliding window of length w. As a result, a binary sequence of length L is transformed into a sequence of $L - w + 1$ w-bit integers (or words), taking some of the 2^w possible values. The probability distribution



(a) $\hat{\lambda} = 0.138891$ (b) $\hat{\lambda} = 0.409249$

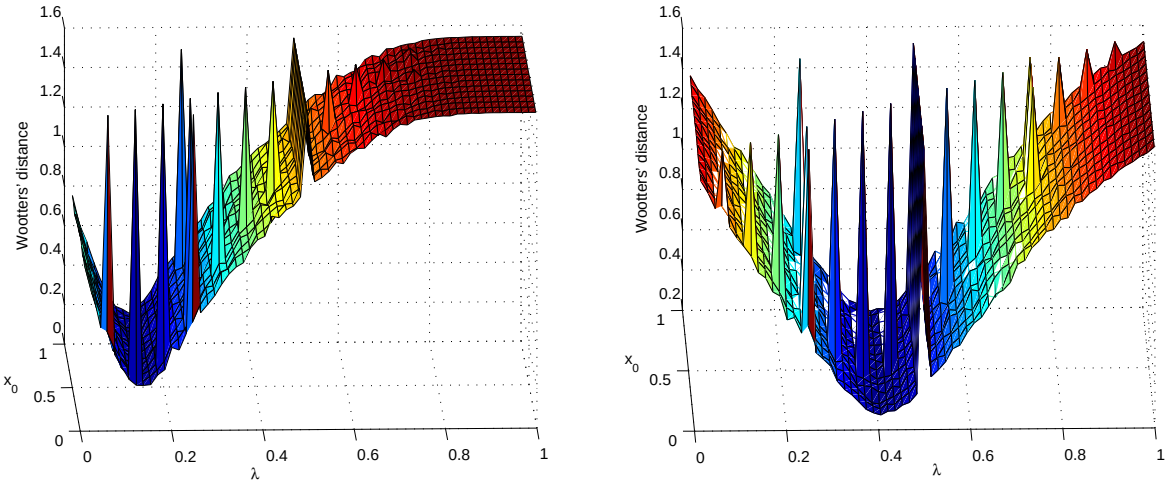
Figure 3: Wootters' distance of the skew tent map with respect to the logistic map for $N = 10^4$ and $w = 10$.

associated to $\mathbf{B}_L(f_\lambda, x_0)$ is determined by counting the number of occurrences of each word and dividing the result by $L - w + 1$.

In the case under consideration, the sequence $\mathbf{B}_{m+N}(f_\lambda, x_0) = \mathbf{B}^{\text{init}} \| \mathbf{B}^{\text{ks}}$ generated at the transmitter, is not accessible to the cryptanalyst. Indeed, as explained in Sec. 3, a chosen-plaintext attack with $\mathbf{P} = \mathbf{0}$ returns $\mathbf{B}^{\text{shuffled}} = \pi(\mathbf{B}_{m+N}(f_\lambda, x_0))$ rather than $\mathbf{B}_{m+N}(f_\lambda, x_0)$, which amounts to the presence of *noise* in the calculated probability distribution. Therefore, the parameters N and w must be selected to guarantee a small value of Wootters' distance between the probability distributions obtained from $\mathbf{B}_{m+N}(f_\lambda, x_0)$ and the one derived from $\mathbf{B}^{\text{shuffled}}$. From this point of view, it is convenient to have a large value of N and a small value of w. On the other hand, the value of w should not be very small, since the entropy of the probability distribution must be as close as possible to the entropy of the underlying chaotic system to achieve an accurate reconstruction of the dynamics involved. Our experience shows that the choice $w \gtrsim 10$ and $N \geq 10^4$ implies a drastic reduction of the noise induced by the shuffling process.

Wootters' distance can be used, for example, to estimate the control parameter of the logistic map. This task is carried out by computing Wootters' distance from the symbolic sequence $\mathbf{B}^{\text{shuffled}}$ (generated with an unknown value $\hat{\lambda}$ of the control parameter) to the symbolic sequences generated with λ ranging in an interval. These distances are computed in Fig. 2 for two values of $\hat{\lambda}$ with $N = 10^4$ and $w = 10$; the corresponding symbolic sequences were generated with different initial conditions. Figure 2 shows that around the right value of λ there exists a basin of attraction, which leads immediately to an estimation of $\hat{\lambda}$. Furthermore, the basin of attraction is always easily observed independently of the shuffling procedure, as it has been verified through different simulations and random interleaving of $\mathbf{B}^{\text{init}}$ and $\mathbf{B}^{\text{ks}}$.

If we consider now a symbolic sequence of the tent map with control parameter $\hat{\lambda}$, then Wootters' distance to the logistic map produces a picture with no basin of attraction (see Fig. 3(a), where the Wootters' distance is always upper 0.9) or with a basin of attraction around $\lambda = 4$ (see Fig. 3(b)). In this case, we conclude that the chosen map is the logistic map with $\hat{\lambda} = 4$, or the tent map with an unknown value for the control parameter. A further analysis of Wootters' distance to the tent map makes possible to discard the logistic map in this situation. Figure 4 depicts Wootters' distance to the tent map when $\mathbf{B}^{\text{shuffled}}$ is generated using the tent map with two different values for $\hat{\lambda}$. Again, it is possible to discern a basin of attraction around $\hat{\lambda}$, which has been verified for different random configurations of the interleaving of $\mathbf{B}^{\text{init}}$ and $\mathbf{B}^{\text{ks}}$. Nevertheless, there is an especial situation where it is impossible to distinguish the logistic map from the tent map. It occurs for the logistic map with $\lambda = 4$ and the skew tent map with $\lambda = 0.5$. In this situation, both maps are topological conjugate [11, p. 68] and the Wootters' distance of the logistic map with respect to the tent



(a) $\hat{\lambda} = 0.138891$ (b) $\hat{\lambda} = 0.409249$

Figure 4: Wootters' distance of the skew tent map with respect to the skew tent map for $N = 10^4$ and $w = 10$.

map shows a basin of attraction around $\lambda = 0.5$. Nevertheless, from a practical point of view it is possible to discern between both maps even when there exists topological conjugacy, since when working with finite precision arithmetics the symmetric tent map possesses a "digitally stable" [1] fixed point at $x = 0$. The effects of digital degradation for the symmetric tent map can be observed in Fig. 4. Indeed, digital degradation is the reason why Wootters' distance with respect to the tent map always shows peaks at $\lambda = 0.5$. However, if the Wootters' distance of the logistic map with $\lambda = 4$ to the skew tent map is calculated, we can observed that it appears a peak instead of a basin of attraction (see Fig. 5(a)), which can be used to distinguish the logistic map from the skew tent map when the theoretical condition of topological conjugacy is satisfied. As a matter of fact, digital degradation causes a dependency of the shape of Wootters' distance with respect to the initial condition of the given symbolic sequence, and also with respect to the quantization steps used in its computation. Future work will be focused on the further and thoroughly examination of that dependence.

We conclude that Wootters' distance between a given symbolic sequence and a large set of symbolic sequences of the logistic map, can discriminate which chaotic map has been used in the encryption procedure. Furthermore, Wootters' distance leads to an estimation of the control parameter of both the logistic map and tent map, which can be further improved as it is shown next.

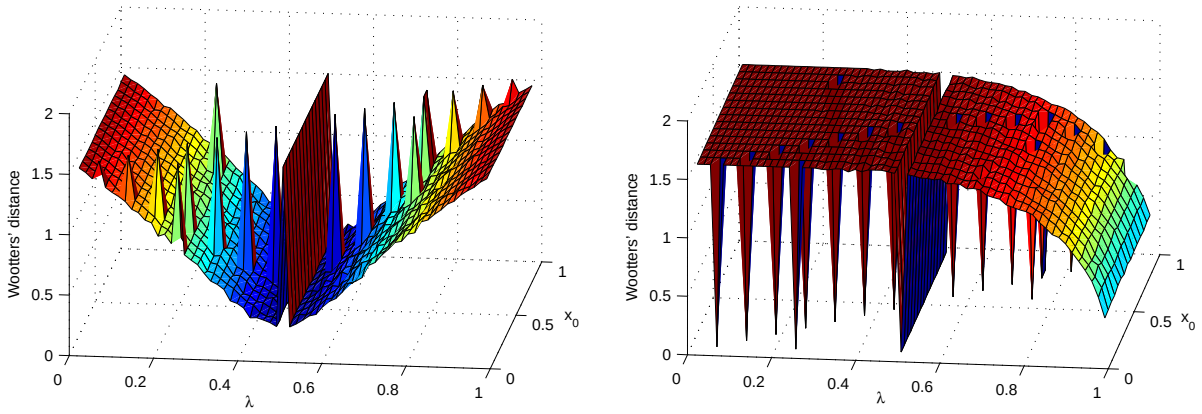
4.2. Estimation of the control parameter from symbolic sequences

As mentioned above, the method to estimate the control parameter from symbolic sequences depends on the underlying chaotic map. In the case of the logistic map, the critical point does not depend on the control parameter, whereas the control parameter determines the critical point for the tent map. This explains the need for different estimation methods.

4.2.1. Control parameter estimation for the logistic map

A method to estimate the control parameter of unimodal maps with a *fixed* critical point using symbolic sequences, can be easily derived from the results of [7], as shown in [9], and applied to cryptanalysis in [13, 14, 15]. As it was mentioned above, in [7] it is proved that for a certain family of unimodal maps $\mathcal{F}$ which includes both the logistic map and the tent map, it is possible to assign a linear order to their

[1] The term "digitally stable" means that the fixed point is stable under finite computing precision. That is, any chaotic orbit will finally lead to $x = 0$ after a limited number of iterations. The number of iterations has an upper bound determined by the finite precision. Some discussions on this phenomenon with floating-point arithmetic can be found in [12].



(a) (b)

Figure 5: Wootters' distance of the (a) logistic map for $\hat{\lambda} = 4$ and $x_0 = 0.593563$; and (b) the skew tent map for $\hat{\lambda} = 0.5$ and $x_0 = 0.213988$ respect to the skew tent map for $N = 10^4$ and $w = 10$. The quantization steps of x- and y-axis are 2^{-6} and 2^{-4} respectively.

symbolic sequences, denoted by $\leq$, that preserves the order of the corresponding initial conditions in $\mathbb{R}$. To be specific, if $I \subset \mathbb{R}$ is a closed interval and $f : I \to I$ belongs to $\mathcal{F}$, then (i) $\mathbf{B}_L(f, x_1) < \mathbf{B}_L(f, x_2)$ implies $x_1 < x_2$, and (ii) $x_1 < x_2$ implies $\mathbf{B}_L(f, x_1) \leq \mathbf{B}_L(f, x_2)$, where $x_1, x_2 \in I$. In particular, for the logistic and tent maps it follows:

(A) $\mathbf{B}_L(f_\lambda, x) \leq \mathbf{B}_L(f_\lambda, f_\lambda(x_c)), \forall x \in [0, 1]$, eventually after a transient orbit in the case of the logistic map.

(B) If $\lambda_1 < \lambda_2$, then $\mathbf{B}_L(f_{\lambda_1}, f_{\lambda_1}(x_c)) \leq \mathbf{B}_L(f_{\lambda_2}, f_{\lambda_2}(x_c))$, since the critical value $f_\lambda(x_c)$ is a non-decreasing, monotone function of λ.

The estimation of the control parameter of the logistic map f_λ is based on (A) and (B), and it proceeds in two stages.

1. Search for the maximum binary sequence of length $l \leq L$ contained in $\mathbf{B}_L(f_\lambda, x_0)$.
2. Use the maximum binary sequence and the monotonic relation between $\mathbf{B}_l(f_\lambda, f_\lambda(x_c))$ and λ, to get an estimation of λ through a binary search procedure [9].

In the scheme defined in [2], $L = m + N$ and, as mentioned before, a chosen-plaintext attack with $\mathbf{P} = \mathbf{0}$ returns $\mathbf{B}^{\text{shuffled}} = \pi(\mathbf{B}_{m+N}(f_\lambda, x_0))$ instead of $\mathbf{B}_{m+N}(f_\lambda, x_0) = \mathbf{B}^{\text{init}} \| \mathbf{B}^{\text{ks}}$. This problem can be overcome by considering not only the maximum binary sequence of length $l \leq m + N$ in $\mathbf{B}_{m+N}(f_\lambda, x_0)$, but the set of the, say, Q greatest sequences of length l. If the interleaving of $\mathbf{B}^{\text{init}}$ in $\mathbf{B}^{\text{ks}}$ is done randomly, it was verified experimentally that for Q large enough, the set of the Q greatest sequences of length l always includes $\mathbf{B}_l^{\max}(f_\lambda, x_0)$ or a good estimation of it, $\mathbf{B}_l^{\max}(f_\lambda, x_0)$ being the maximum sequence obtained from $\mathbf{B}_{m+N}(f_\lambda, x_0)$. In [15] it is pointed out that a good estimation of λ requires values of Q over $10^6 \approx 2^{20}$ (a typical number in actual chosen-plaintext attacks); the estimation error lies then below 10^{-8} (see Fig.1 in [15]). In our case, this estimation is also degraded by the fact that the method is applied on approximated values of $\mathbf{B}_l^{\max}(f_\lambda, x_0)$. Needless to say, an estimation of λ amounts to reducing the key space, and this compromises the security of the cipher. All in all this analysis underlines the critical role of the shuffler in the encryption scheme of [2].

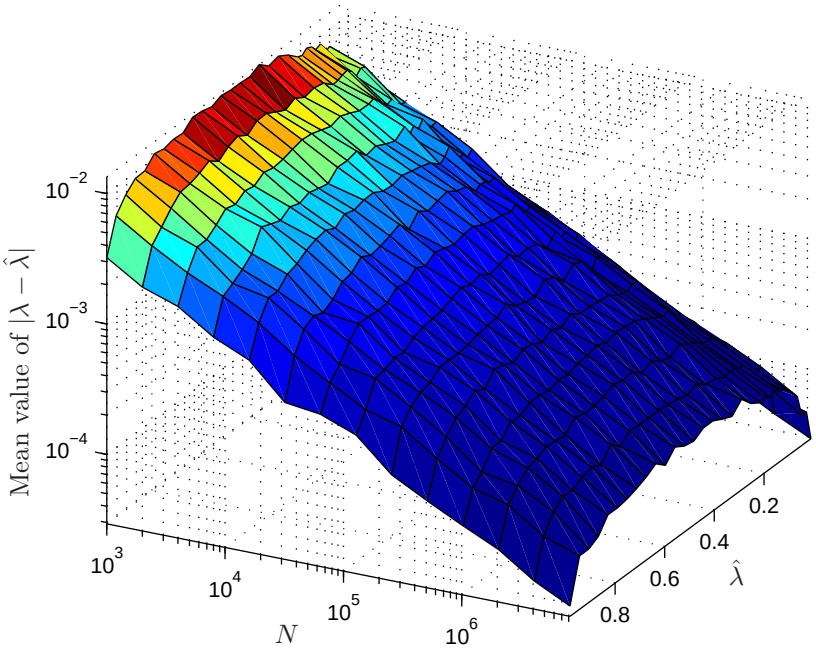


Figure 6: Error in the estimation of the control parameter of the tent map from the ratio between 1-bits and 0-bits in $\mathbf{B}^{\text{shuffled}}$.

4.2.2. Control parameter estimation for the tent map

The method described in the previous section does not apply to the tent map because its critical point depends on the control parameter: $x_c = \lambda$. In this case, we can resort to the analysis of the ratio between 1-bits and 0-bits in symbolic sequences of the tent map. As it was emphasized in Sec. 3.1, this ratio is equal to $R = \frac{1-\lambda}{\lambda}$, hence it can be used to estimate of the control parameter. Moreover, the number of 1-bits and 0-bits is not modified by the shuffling procedure, so the estimation of the control parameter can be performed on $\mathbf{B}^{\text{shuffled}}$ instead. Fig. 6 shows the error in the recovery of the control parameter from the ratio R obtained with different $\mathbf{B}^{\text{shuffled}}$ and λ ranging in $(0, 1)$. The estimation error decreases as the length of the plaintext N increases, but a perfect recovery of λ requires not only large values of N but also extended-precision arithmetic libraries. Indeed, when implementing the cryptosystem, the shortcomings of finite precision arithmetic and finite statistical sampling causes a deviation of the computed value of R from its theoretical value, which further entails a residual error in the estimation of λ. It was experimentally verified that this residual error is around 10^{-4}, the numerical simulations being carried out with double-precision floating-point arithmetic. In any case, the estimation of the control parameter of both the logistic map and tent map, implies a severe reduction of the key space that must be taken into account when designing the cipher.

4.3. Digital degradation

A main characteristic of stream ciphers is that the keystreams must have a very long period. In the context under examination, the period of the keystreams depends on the periodic behavior of the symbolic sequences of the logistic map and the tent map. It is well known that any finite-precision orbit, hence any symbolic sequence of chaotic map, is periodic. This problem is especially important in the case of the *symmetric* tent map. In particular, for $\lambda = 0.5$ the origin is an attractive fix point for all orbits, and this represents a complete degradation of the random properties of the corresponding keystreams. Therefore,

the recommendations given in [6] must be taken into account in order to avoid the consequences of the dynamical degradation.

5. Conclusions

In this paper we have analyzed a recent stream cipher that is built on the symbolic sequences of the (parametric) logistic and tent maps. We conclude that this cipher is insecure since a chosen-plaintext attack makes possible to estimate the control parameter of the underlying chaotic map, based on a "noisy" version of the keystream. This estimation can be done with an approximate error that goes from 10^{-4} (tent map) to 10^{-8} (logistic map), what amounts to a strong reduction of the key space. More generally, the results of this paper and of [16] hint to the fact that symbolic sequences of unimodal maps are insecure when used as keystreams.

Acknowledgments

The work described in this paper was supported by *Ministerio de Educación y Ciencia of Spain*, research grant SEG2004-02418, *CDTI, Ministerio de Industria, Turismo y Comercio of Spain* in collaboration with Telefónica I+D, Project SEGUR@ with reference CENIT-2007 2004, *CDTI, Ministerio de Industria, Turismo y Comercio of Spain* in collaboration with SAC, project HESPERIA (CENIT 2006-2009), and *Ministerio de Ciencia e Innovación of Spain*, project CUCO (MTM2008-02194). Shujun Li was supported by a fellowship from the Zukunftskolleg of the Universität Konstanz, Germany, which is part of the "Exzellenzinitiative" Program of the DFG (German Research Foundation).

References

[1] T. Stojanovski, L. Kocarev, Chaos-based random number generators-part I: analysis, Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on 48 (3) (2001) 281–288. doi:10.1109/81.915385.

[2] A. P. Kurian, S. Puthusserypady, Self-synchronizing chaotic stream ciphers, Signal Processing 88 (2008) 2442–2452.

[3] T. Stojanovski, L. Kocarev, R. Harris, Applications of symbolic dynamics in chaos synchronization, IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 44 (10) (1997) 1014–1018.

[4] G. Alvarez, S. Li, Some basic cryptographic requirements for chaos-based cryptosystems, Int. J. Bifurc. Chaos 16 (8) (2006) 2129–2151.

[5] D. Arroyo, G. Alvarez, V. Fernandez, On the inadequacy of the logistic map for cryptographic applications, in: L. Hernandez, A. Martin (Eds.), X Reunión Española sobre Criptología y Seguridad de la Información (X RECSI), Universidad de Salamanca, Salamanca, Spain, 2008, pp. 77–82, (ISBN 978–84–691–5158–7).

[6] S. Li, G. Chen, X. Mou, On the dynamical degradation of digital piecewise linear chaotic maps, International Journal on Bifurcation and Chaos 15 (10) (2005) 3119–3151.

[7] N. Metropolis, M. Stein, P. Stein, On the limit sets for transformations on the unit interval, Journal of Combinatorial Theory (A) 15 (1973) 25–44.

[8] G. Alvarez, D. Arroyo, J. Nunez, Application of gray code to the cryptanalysis of chaotic cryptosystems, in: 3rd International IEEE Scientific Conference on Physics and Control (PhysCon'2007, 3rd - 7th, September 2007, Potsdam, Germany), IEEE IPACS, Potsdam, Germany, 2007.
URL http://lib.physcon.ru/?item=1358

[9] X. Wu, H. Hu, B. Zhang, Parameter estimation only from the symbolic sequences generated by chaos system, Chaos, Solitons and Fractals 22 (2004) 359–366.

[10] A. P. Majtey, P. W. Lamberti, M. T. Martin, A. Plastino, Wootters' distance revisited: a new distinguishability criterium, Eur. Phys. J. D 32 (2005) 413–419. doi:10.1140/epjd/e2005-00005-1.

[11] B.-L. Hao, W.-M. Zheng, Applied symbolic dynamics and chaos, Vol. 7, Directions in Chaos, 1998.

[12] S. Li, When chaos meets computers, last revised in December 2005 (2004).
URL http://arxiv.org/abs/nlin.CD/0405038

[13] G. Alvarez, F. Montoya, M. Romera, G. Pastor, Cryptanalysis of an ergodic chaotic cipher, Physics Letters A 311 (2003) 172–179.

[14] K. Wang, W. Pei, L. Zou, A. Song, Z. He, On the security of 3D cat map based on symmetric image encryption scheme, Physics Letters A 343 (2005) 432–439.

[15] D. Arroyo, G. Alvarez, S. Li, C. Li, V. Fernandez, Cryptanalysis of a new chaotic cryptosystem based on ergodicity, International Journal of Modern Physics B (2008) Accepted in November.
URL http://arxiv.org/abs/0806.3183

[16] D. Arroyo, G. Alvarez, J. M. Amigó, Estimation of the control parameter from symbolic sequences: Unimodal maps with variable critical p
(2008).
URL http://arxiv.org/abs/0812.2331