International Journal of Cryptology Research 1 (1): 93-101 (2009)

Conditions for Counter Measures Against One Time Pad Attack on Baptista Type Chaotic Cryptosystem

[1]M.R.K.Ariffin and [2]M.S.M.Noorani
[1]Institute for Mathematical Research
Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]School of Mathematical Sciences,
Faculty of Science and Technology, Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, Malaysia
Email: mdrezalka@gmail.com

ABSTRACT

In 1998, M.S. Baptista proposed a chaotic cryptosystem using the ergodicity property of the simple low-dimensional and chaotic logistic equation $X_{n+1} = bX_n(1 - X_n)$ where X_0 and b are the secret keys. This cryptosystem has the ability to produce various ciphers responding to the same message input. Since then, many cryptosystems based on Baptista's work has been proposed. However, over the years research has shown that this cryptosystem is predictable and vulnerable to attacks and is widely discussed. Among the weaknesses are the non-uniform distribution of ciphertexts and succumbing to the one-time pad attack (a type of chosen plaintext attack). The one-time pad attack which was constructed by Alvarez (2003) proved that the ergodic cipher put forward by Baptista behaves as a one-time pad which reuses its key, and as a result, is easy to break. The method of attack is based on the symbolic dynamics of one dimensional quadratic map. The focus of our research is to overcome the one-time pad attack. As pointed out by Alvarez, obtaining the one-time pad is as good as knowing the key (i.e. X_0 and b), making the system 100% vulnerable. We give a formal treatment for the one-time pad attack. We derive definitions and give mathematical explanations for this phenomenon. Finally, we give a theorem, if satisfied by a "counter measure" method, would result in this cryptosystem being invulnerable against the one-time pad attack.

Keywords: Chaotic cryptosystems, Ergodicity, Cryptanalysis, Logistic map

INTRODUCTION

M.S Baptista proposed in 1998 a new cryptosystem based on the property of ergodicity of chaotic systems (i.e. the eventual visit of the trajectory to all partitions in the phase space as the number of iteration grows). This mathematical property induces this cryptosystem type with an ability to produce various ciphers responding to the same message input. In other words, this type of cryptosystem is a dynamic cryptosystem due to mathematical considerations and not due to computer programming methods. His cryptosystem utilizes the logistic map

$$x(n+1) = bx(n)(1 - x(n)) \tag{1}$$

where $x(i) \in [0,1]$ and the parameter b is chosen such that eq (1) behaves chaotically.

Motivated by the interest in chaotic cryptosystem, and by Baptosta's ergodic cipher, numerous algorithms based on variations of Baptista's have been proposed. In 2003 Alvarez presented 4 different cryptanalytic attacks after a complete study of Baptista' algorithms: one-time pad attacks, entropy attacks and key recovery attacks, comprised of parameter and initial point estimation. In 2004 Alvarez cryptanalyzed a variation of Baptista by Wong (2002).

Among the 4 types of attacks, the one-time pad attack is the most efficient. It is a type of known plaintext attack. Without the knowledge of the secret parameters and the initial condition, the one-time pad attack can decrypt any message encrypted via this method. The one-time pad attack capitalizes on the fact that the ergodic cipher put forward by Baptista behaves as a one-time pad which reuses its key, and as a result, easy to break.

ONE-TIME PAD ATTACKS

The Baptista ergodic cipher behaves as a modified version of the one-time pad. A one time pad is perfectly secure under certain conditions. It is crucial to the security of the one-time pad that the key be as long as the message and never be reused, thus preventing 2 different messages encrypted with the same portion of the key being intercepted or generated by an attacker.

Alvarez proved that the one-time pad attack capitalizes on the fact that the ergodic cipher put forward by Baptista behaves as a one-time pad which reuses its key, and as a result, easy to break. Alvarez's attack is based on symbolic dynamics of 1-D quadratic maps, such as the logistic map given by (1).

Since it is a known plaintext attack, a few pairs of cipher and plain texts are requested. Every pair of cipher and plain text allows recovering exactly 50% of the one-time pad. After analyzing one pair, the probability of finding any correct symbol of the one-time pad is exactly $1-\frac{1}{2}=0.5$. If a second pair is analyzed, and assuming that plain texts are perfectly random and mutually independent, the probability of finding any symbol of the one-time pad is $1-\frac{1}{2^2}=0.75$. With p different plaintext messages, the

probability of finding a correct symbol approaches $1-\frac{1}{2^p}$. This is the maximum rate of convergence to the one-time pad for small values of p, independent of the key length.

Alvarez used the following illustration. Let us use assume that unknown to the cryptanalyst, the keys are $X_0 = 0.232323$ and $b = 3.78$, using the interval [0.2,0.8]. Let us use a 4-symbol source $S_4 = \{s_1, s_2, s_3, s_4\}$. Under these assumptions, in a chosen plaintext attack scenario, we request the ciphertext of messages consisting of all their symbols set to s_1, s_2, s_3 and s_4, respectively.

4-symbol alphabet	Associated sites
s_4	$[0.65,0.80)$
s_3	$[0.50,0.65)$
s_2	$[0.35,0.50)$
s_1	$[0.2,0.35)$

We begin by requesting the ciphertext for a plaintext consisting of only $s_1 : P = (s_1, s_1, s_1, s_1, s_1, s_1, \ldots)$. The corresponding ciphertext is $E_1 = (5,9,5,3,4,5, \ldots)$. Examining the ciphertext, we know for sure that the 6th symbol in the one-time pad is s_1, that the 14th symbol is another s_1, and the 20th, and the 23rd and so on. After considering the whole message, we get the following partial sequence for the one-time pad $O = xxxxxs_1xxxxxxxxs_1xxxxs_1xxs_1xxxs_1xxxxs_1\ldots$. Since we chose the interval [0.2,0.8] instead of the whole attractor, the letters marked x could correspond to either an iteration below the lower bound 0.2 or beyond the upper bound 0.8 or it could also correspond to the other symbols from the 4-symbol source.

For s_2, the corresponding ciphertext is $E_2 = (8,2,2,5,7,5,\ldots)$. We can complete more gaps in the sequence $O = xxxxxs_1xxs_2xs_2xs_2xs_1xxs_2xs_1xxs_1xs_2xs_1xxs_2xs_1x\ldots.$

For $S_{3,}E_3 = (3,3,14,7,\ldots)$ and $O = xxxs_3xs_1s_3xs_2xs_2xs_2xs_1xxs_2xs_1s_3xs_1xs_2xs_1s_3xs_2xs_1x\ldots.$ Lastly for s_4, the ciphertext is given by $E_4 = (1,14,17,1,\ldots)$, and we are able to construct $O = xs_4xs_3xs_1s_3xs_2xs_2xs_2xs_2s_4xs_2xs_1s_3xs_1xs_2xs_1s_3xs_2x_1s_1s_4\ldots.$

Hence, the one-time pad is constructed. The symbol x correspond to points outside the boundaries and cannot be used for encryption. It is important to note that knowing the one-time pad generated by a certain key (X_0 and b) is entirely equivalent to knowing the key.

As an example, let us first construct the ε-intervals with its corresponding source. Since $n = 4$, we have $\varepsilon = 0.15$, and s_1 is associated with the interval $[0.2,0.35)$, s_2 with the interval $[0.35,0.5)$, s_3 with the interval $[0.5,0.65)$ and finally s_4 with the interval $[0.65,0.8)$. If given the following ciphertext $C^* = (1,4,3,2,2,3,5)$, it is easily derived from the one-time pad that the corresponding plaintext is $P = s_4, s_1, s_2, s_2, s_2, s_4, s_3$.

CONDITIONS FOR COUNTER MEASURES AGAINST THE ONE-TIME PAD ATTACK

In this section we give a formal treatment for the one-time pad attack. We derive definitions and give mathematical explanations for this phenomenon. Finally, we give a theorem, if satisfied by a **"counter measure" method**, will relieve this cryptosystem which has the ability to produce various ciphers responding to the same message input from the one-time pad attack.

Definition 1 (Associated sites)

Let $A = (a,b) \subset [0,1]$. For an alphabet of *k*-characters, the associated site for each member of the alphabet s_i $i = 1,2,\ldots,k$ is an interval $A_i \subset A$, where $s_i \in A_i$.

$A_i = (a + (i-1)\varepsilon, a + i\varepsilon)$ where $\varepsilon = \dfrac{b-a}{k}$.

Definition 2

Let $E_i = \{s_{i,m}\}_{m=1}^{n}$ be an ordered finite sequence that represents the number of iterations a chaotic map on a pre-defined initial condition x_0, needs to arrive at an associated site of a designated alphabet s_i.

From definition 2, let

$$\alpha_{i,1} = \sum_{m=1}^{1} s_{i,m}$$
$$\alpha_{i,2} = \sum_{m=1}^{2} s_{i,m}$$
$$\vdots$$
$$\alpha_{i,n} = \sum_{m=1}^{n} s_{i,m}$$

Definition 3

Let E_i^* be the cumulative set where $E_i^* = \{\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n}\}$.

Motivated by observation and evidence provided by Alvarez, we derive the following definition.

Definition 4

A chaotic map on a predefined initial condition x_0, behaves as a OTP that reuses its key when $\bigcap_{i=1}^{j} E_i^* = \varnothing$.

Motivated from computational evidence, for the logistic map, the following statement is derived.

Statement

If a finite sequence of identical s_i's is substituted arbitrarily by a sequence of s_r's where $r \in \{1,2,\ldots,k\}$ then $\bigcap_{i=1}^{j} E_i^* \neq \varnothing$. That is, the chaotic map that generates each E_i^* no longer behaves as an OTP that reuses its key.

Example 1:

The following substitution is done prior to giving the ciphertext for a plaintext consisting of identical elements. The first row represents the plaintext. The second row represents the substitution for the list plaintext.

s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1	s_1
s_4	s_2	s_4	s_3	s_1	s_4	s_2	s_1	s_1	s_1	s_2	s_2	s_1	s_4	s_1	s_1	s_3	s_2	s_1	s_3

s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2	s_2
s_2	s_1	s_2	s_3	s_3	s_1	s_2	s_4	s_2	s_3	s_4	s_3	s_1	s_3	s_3	s_4	s_4	s_3	s_1	s_1

s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3	s_3
s_3	s_4	s_4	s_3	s_2	s_4	s_3	s_1	s_1	s_3	s_2	s_3	s_2	s_4	s_4	s_2	s_3	s_3	s_4	s_2
s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4	s_4
s_2	s_1	s_3	s_2	s_1	s_1	s_2	s_3	s_2	s_1	s_3	s_1	s_1	s_3	s_1	s_3	s_1	s_2	s_1	s_3

We will use the one-dimensional logistic map with the initial condition $X_0 = 0.232323$ and $b = 3.78$. We begin by requesting the ciphertext for a plaintext consisting of only $s_1 : P = (s_1, s_1, s_1, s_1, s_1, s_1, \ldots)$. The corresponding ciphertext $E_1 = (2,6,1,6,4,2,1,4,21,5,\ldots)$.For s_2, the corresponding ciphertext is $E_2 = (8,6,3,3,7,20,3,1,2,2,\ldots)$. For s_3, $E_3 = (1,1,2,2,2,1,6,4,7,1,2,\ldots)$ and lastly for s_4, the ciphertext is given by $E_4 = (8,6,1,2,2,7,3,6,4,8,\ldots)$.

We are able to construct the following partial one-time pad:

$$O = x\ s_3\ s_1s_3\ x\ s_3x\ s_3x\ s_1s_2s_3s_4\ s_1s_3\ xxxx\ s_2s_4\ s_1s_3s_4\ x\ s_2s_4\ s_1s_3s_4\ s_2\ s_1\ s_1$$
$$xxx\ s_1s_3s_4\ s_2s_3\ x\ s_3s_4\ xxxxx\ s_4\ xxx\ s_4\ xxxxxxx\ s_1s_2s_4\ xx\ s_2\ s_2\ s_1\ s_2x\ s_2\ \ldots$$

From the above partial one-time pad, it can observe that there are $4^1X3^4X2^6$=20,736 possibilities to choose from. The longer the one-time pad or the more alphabets that are involved in the symbol source, one cannot rule out the posiibility of having more than 2^{1024} possibilities.

The following theorem explains this phenomenon in a mathematical setting.

Theorem 1

The one-time pad attack on Baptista type chaotic cryptosystem fails if and only if $\bigcap_{i=1}^{j} E_i^* \neq \varnothing$.

Proof:

$(\Rightarrow)$

First, let us conduct the OTP attack by requesting the ciphertexts of each alphabet. We will be given the following sets:

$$E_1 = \{s_{1,1}, s_{1,2}, \ldots, s_{1,n}\}$$
$$\vdots$$
$$E_i = \{s_{i,1}, s_{i,2}, \ldots, s_{i,n}\}$$

where $i = 1,2,\ldots,k$.

We will then obtain the cumulative sets;

$$E_i^* = \{\alpha_{1,1}, \alpha_{1,2}, \ldots, \alpha_{1,n}\}$$
$$\vdots$$
$$E_i^* = \{\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n}\}$$

where $i = 1,2,\ldots,k$.

The OTP attack fails when there exists $\alpha_{j,l} \in E_j^*$ such that for some $p = 1,2,\ldots,k$ and $p \neq j$, $\alpha_{j,l} \in E_p^*$. That is, $\bigcap_{i=1}^{j} E_i^* \neq \varnothing$.

$(\Leftarrow)$

If $\bigcap_{i=1}^{j} E_i^* \neq \varnothing$, there exists $\alpha_{j,l} \in E_j^*$ such that for some $p = 1,2,\ldots,k$ and $p \neq j$, $\alpha_{j,l} \in E_p^*$. For a OTP attack on a Baptista type chaotic cryptosystem to be successful, elements from distinct sets E_i^* where $i = 1,2,\ldots,k$, must not overlap (hence the term OTP). Thus, the OTP attack fails.

CONCLUSION

From results provided in section 3.0, it is obvious that for any cryptographer intending to enhance the Baptista type chaotic cryptosystem such that it is invulnerable from the one-time pad attack, one has to induce the original algorithm with elements such that $\bigcap_{i=1}^{j} E_i^* \neq \varnothing$. And if the above example is to be taken into consideration, a reversible substitution method must be designed.

ACKNOWLEDGEMENTS

This research was supported by the Science Fund, Ministry of Science, Technology and Innovation, Malaysia under grant number 04-01-02-SF0177.

REFERENCES

[1] C.E. Shannon. 1949. *Bell Systems Tech. J.* ,28:656.

[2] G. Grassi, S. Mascolo. 1998. *Electron. Lett.* ,34:1844.

[3] Y.H. Chu, S.Chang. 1999. *Electron. Lett.* ,35:271.

[4] T.Yang Tao, C.W.Wu, L.O.Chua. 1997. *IEEE Trans. CASI* , 44:469

[5] M.S.Baptista. 1998. *Phys. Lett. A* , 240:50.

[6] E. Alvarez, A.Fernandez, P.Garcia, J.Jimenez, A.Marcano. 1999. *Phys. Lett. A*, 3 :373.

[7] W.Wong, L.Lee, K.Wong. 2001. *Comput. Phys.Commun.* 138:234.

[8] K.Wong. 2002. *Phys. Lett. A* 298:238.

[9] G.Alvarez, F.Montoya,M.Romera,G.Pastor. 2003. *Phys. Lett. A* 311:172.

[10] G.Alvarez, F.Montoya,M.Romera,G.Pastor. 2004. *Phys. Lett. A* 326:211.