THE ANNALS OF "DUNĂREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2012, VOL. 35, NO. 1, ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

MULTIPLE BITSTREAMS GENERATION USING CHAOTIC SEQUENCES

Adrian-Viorel DIACONU – PhD candidate[1], Assistant Professor[2]

[1]University Politehnica of Bucharest, ETTI Faculty, Bucharest, Romania
[2]Lumina - The University of South-East Europe, IT&C Dept., Bucharest, Romania
(e-mail: adrian.diaconu@lumina.org)

Abstract: This article presents a method to transform a chaotic sequence of real numbers into multiple bitstreams, using multi-level discretization. Although the original intent was to achieve a higher/multiple flow of bits, much faster (i.e. in terms of computational time), it was subsequently conjugated with the one of preserving statistical properties of newly formed bitstreams (i.e. to have the ability of keeping their recommendations for usage within secure cryptographic application and not only, e.g. stochastic computing and/or generation of simulation data for the study of traffic in networks, whether they are WSNs, ATM or Telecoms). Thus, CrypTool, VRA and NIST battery of statistical tests were used in order to present an analysis of the randomness of the bitstreams obtained by applying the proposed discretization method. Theoretical and practical arguments, rounded by good statistical results, confirm viability of the proposed method and recommend it in generation of multiple bitstreams that will be used for secure cryptographic applications.

Keywords: chaotic dynamical system, randomness testing, multi-level discretization, VRA, NIST.

INTRODUCTION

In last two decades, interesting relationships between chaos and cryptography have been developed, many properties of chaotic systems (e.g. mixing properties, ergodicity, sensitivity to initial conditions or system's parameters, structural complexity and deterministic dynamics) being considered analogous to confusion, diffusion with small change in plaintext or secret key, diffusion with a small change within one block of the plaintext, deterministic pseudo-randomness and / or algorithmic complexity of traditional cryptosystems. As a result, several chaos-based cryptosystems have been put forward since 1990. Although chaos implies unpredictable time behavior of system, its dynamics (whose evolution seems to be true random) can be expressed by one or more deterministic rules. This is one of the main properties of chaotic dynamical systems which have encouraged the idea to design new pseudo-random number generators and also to develop some robust encryption schemes based on chaotic sequences (Alvarez and Li, 2006; Li, *et al.*, 2001; Kocarev, *et al.*, 1998; Li, *et al.*, 2003).

Cryptosystems are not the only applications where chaos-based PRNG found their place; e.g. WSNs are highly vulnerable to the failure of BS (i.e. adversaries can easily render WSNs useless by launching remote software / physical based attacks on the BS), leading to few research works, as the one conducted by Jing, *et al.* (2005), which address the problem of defending BSs against physical attacks (i.e. by concealing BSs'

This paper was recommended for publication by Viorel Nicolau

geographic location, using some randomness degrees, introduced in different paths within the WSN).

Obviously, purpose for which chaos-based PRNGs were designed and widely used does not just stop at these applications. This ample range of applications substantiated the present research, on the generation of multiple bitstreams using chaotic sequences.

PROPOSED DISCRETIZATION METHOD

2.1. Chaotic maps and their discretization method

The logistic and tent (1) maps are most widely used maps into designing of brand new digital chaotic cryptosystems (Arroyo, *et al.*, 2008; Alvarez, *et al.*, 2012; Luca, *et al.*, 2009; Şerbănescu, *et al.*, 2008; Li, *et al.*, 2005). Without insisting on theoretical and practical aspects, related to these maps' exploitation, typical discretization method (2) of (1) is presented (Sebesta, 2007).

$$(1) \qquad f_T : [0,1] \to [0,1],\ f_T(x) = r(1 - |1 - 2x|),\ r \in (0,1)$$

$$(2) \qquad b_{c,n} = \begin{cases} 0 & \textit{for} \quad f_T(x_{c,n}) < 0.5 \quad \textit{and} \\ 1 & \textit{for} \quad f_T(x_{c,n}) \geq 0.5 \end{cases}$$

2.2. Proposed multi-level discretization method

For generation of multiple bitstreams, using chaotic sequence, the multi-level discretization method (3) is proposed.

$$(3) \qquad b_{c,n} = \begin{cases} B_1B_0 & \textit{for} \quad f_T(x_{c,n}) \in [0.00, 0.25] \\ B_1B_0 & \textit{for} \quad f_T(x_{c,n}) \in [0.25, 0.50] \\ B_1B_0 & \textit{for} \quad f_T(x_{c,n}) \in [0.50, 0.75] \\ B_1B_0 & \textit{for} \quad f_T(x_{c,n}) \in [0.75, 1.00] \end{cases},\ B_1B_0 \in \{00, 01, 10, 11\}$$

Resulted di-bits are spreaded into two separate files, called bitstream_A.txt (which contains di-bit's first bit) and bitstream_B.txt (containing di-bit's second bit).

3. STATISTICAL TESTING

In order to assess bitstreams' suitability within any cryptographic appl. (i.e. their statistical properties, true randomness), different tools sunch as CrypTool, VRA and NIST were used. Operating methodology, for each of them, and obtained results are presented and discussed in the following subsections.

4.1. CrypTool analysis

From the information theory, by Claude E. Shanon (1948), CrypTool[1] was used to compute frequencies of binary strings composed of *n* characters (i.e. the *n*-grams). For true random strings, is expected that each entry within *n*-gram has the same probability of occurrence, given by (4).

$$(4) \qquad P[\%] = \frac{i}{2^n} \cdot 100$$

where, *i* represents the binary string's length and *n* represents the *n*-gram's order.

n-gram statistics were performed over 100 randomly chosen binary sequences, each sequence of length *i* = 10.000.000 bits, overall results being presented in table 1.Table 1. n-gram reports

n-gram's order		***bitstream***	
		A	***B***
Histogram (*n* = 1)	1	50.3429%	50.3124%
	0	49.6571%	49.6876%
Digram (*n* = 2)	00	24.5009%	24.5581%
	01	25.1562%	25.1294%
	10	25.1562%	25.1295%
	11	25.1866%	25.1829%
Trigram (*n* = 3)	000	11.9561%	12.0423%
	001	12.5448%	12.5157%
	010	12.5714%	12.5559%
	011	12.5848%	12.5735%
	100	12.5448%	12.5158%
	101	12.6114%	12.6137%
	110	12.5847%	12.5736%
	111	12.6018%	12.6093%
4-Gram (*n* = 4)	0000	5.7275%	5.8052%
	0001	6.2286%	6.2370%
	0010	6.2670%	6.2455%
	0011	6.2778%	6.2702%
	0100	6.2666%	6.2627%
	0101	6.3048%	6.2932%
	0110	6.2985%	6.2692%
	0111	6.2863%	6.3043%
	1000	6.2286%	6.2371%
	1001	6.3162%	6.2787%
	1010	6.3044%	6.3104%
	1011	6.3070%	6.3033%
	1100	6.2782%	6.2531%
	1101	6.3065%	6.3205%
	1110	6.2862%	6.3044%
	1111	6.3155%	6.3049%
12-Gram (*n* = 12)	00…00	-	0.0109%
	…	≈ 0.0244%	≈ 0.0244%
	11…11	0.0231%	0.0287%

Table's 1 analysis not only doesn't emphasizes the dominant presence of any sub-strings (i.e. in terms of frequency of use) (Banerjee and Pedersen, 2003; Muise, *et al.*, 2009) but also highlights a uniform system dynamics (i.e. in terms of the time evolution of f_T's trajectories, balanced between all four thresholds). Therefore, positive results obtained at this point guide us to perform the next statistical analysis.

4.2. VRA analysis

RPs (i.e. **R**ecurrence **P**lots) yield very important insights into the time evolution of f_T's trajectories, because typical patterns in RPs are linked to specific

[1] Open-source program offering an innovative visual programming GUI to experiment with cryptographic procedures and to animate their cascades. Last accessed on [01.10.2012]: www.cryptool.com.

system behaviour (Marwan, *et al.*, 2007). Without proper settings of analysis parameters resulted RPs will be images completely devoided of information. To obtain as much information, suitable embedding dimension and adequate time delay must be chosen. Using MI and FNN (i.e. **M**utual **I**nformation, and **F**alse **N**earest **N**eighbours) VRA's toolboxes these parameters can be correctly set to the optimal value (Jorge and Dulce, 2002).

MI and FNN were performed on each bitstream (each having 1.000.000 bits in length, generated with 100 randomly chosen seeds), with the aid of resulting parameters (i.e. embeding dimension $m = 5$ and time delay $d = 3$) each bitstream's RPI being computed, as shown in Fig. no. 1 and Fig. no. 2.

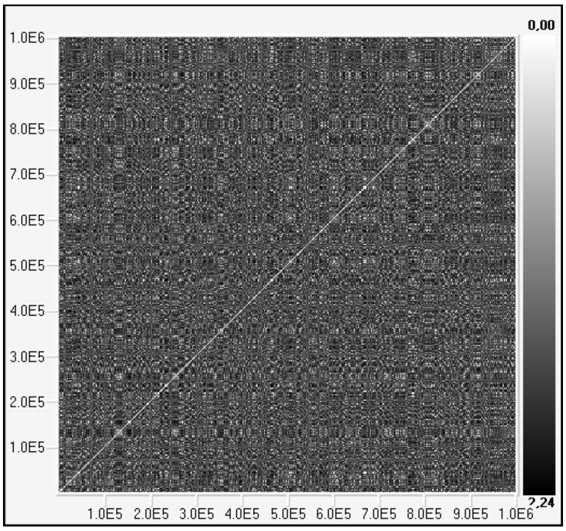

Fig. no. 1 RPI of bitsream A

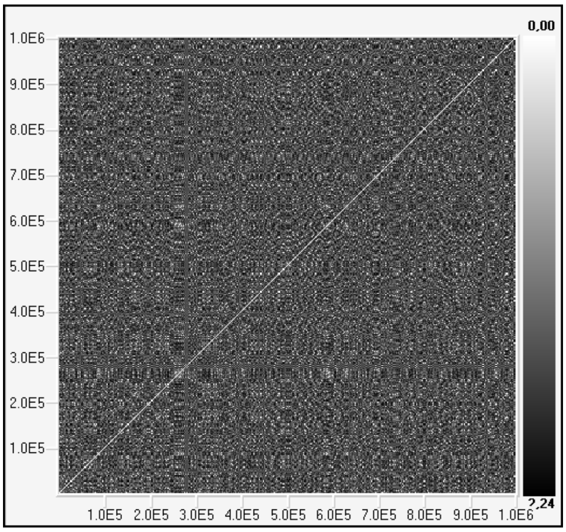

Fig. no. 2 RPI of bitsream B

Lack in clear patterns, in either RPI, indicates that consecutive samples in bitsreams' structure are much far apart and uncorrelated. RPIs' homogeneity along the major diagonal and irregular distribution highlights a stationary, mostly stocastic behaviour (i.e. intrinsically non-deterministic, non-intermitent and sporadic) of the system that has generated the bitstreams and, namely, a true random process (i.e. random binary strings).

Table 2. VRA general statistics

statistics	***bitstream***	
	A	***B***
Mean	0.5034	0.5031
Variance	0.2500	0.2500
Standard deviation	0.5000	0.5000
Skewness	- 0.0137	- 0.0124
Kurtosis	- 1.9998	- 2.0001

VRA Tool provides and other additional general statistics (i.e. mean, standard deviation, percentage of recurrence and determinism, entropy etc.), some of them, the most important ones, being quantified in table 2. It can be noticed that standard deviation and variance have values close to ideal, except of mean (as it was expected, refer to histogram values).

Despite the fact that Skewness has a negative value (i.e. indicating that the tail on the left side of the probability density function is longer than the right side and the bulk of the values lie to the right of the mean) its close to zero value indicates that the values are relatively evenly distributed on both sides of the mean, typically (but not necessarily) implying a symmetric distribution (Doane and Seward, 2011). At the same time, Kurtosis's high level and negative value denotes a platykurtic distribution (i.e. data set with flatter peak around its mean, which causes thin tails within the distribution and low level of data fluctuation) (DeCarlo, 1997).

Good general statistical properties revealed with the aid of VRA (i.e. visually – evaluation of structural properties or through RQA – numerical properties quantification), highlights randomness of bitstreams generated using f_T, in conjuction with the proposed discretization method, allowing advance to NIST statistical testing.

4.3. NIST statistical testing

In order to determine system's security level (i.e. of the system resulting from the implementation of proposed multi-level discretization method, on chaotic sequences of real numbers) against some statistical cryptanalytic attacks (Alvarez, *et al.*, 2003; Patidar and Sud, 2009; Kocarev and Lian, 2011), NIST standard battery of tests (Rukhin, *et al.*, 2010; Kim, *et al.*, 2004) was used, as a tool for statistical analysis of randomness.

Thus, we undergone to the assessment of the results obtained from testing the randomness of the values generated by the orbits of f_T map and, subsequentlly, subjected to discretization process, with four thresholds (i.e. 2-bit encoding of each interval).

First NIST tests were performed on bitsteams computed with the first lexicographically generated encodings of the thresholds (i.e. [0,0.25] coded as 00, [0.25,0.50] as 01, [0.50,0.75] as 10, [0.75,1] as 11). For the numerical experimentations were generated 2000 (i.e. sample size m = 2000) different binary sequences from 100 randomly chosen seeds, each sequence having a length of n = 1.000.000 bits, and computed p-value corresponding to each sequence for all the 17 tests of the NIST suite.

The significance level of each test in NIST is set to 1%, which means that 99% of test samples pass the tests if the random binary sequence is truly random. The acceptance region of the passing ratio is given by (5), where m represents the number of samples tested and $p = 1 - \alpha$ is the probability of passing each test. For m = 2000 and the probability p = 0.99 (i.e. corresponding to the significance level α = 0.01) the confidence interval [0.983, 0.996] was obtained.

$$(5) \qquad \left[p-3\sqrt{\frac{p\cdot(1-p)}{m}}, p+3\sqrt{\frac{p\cdot(1-p)}{m}} \right]$$

In tables 3 and 4 the results, obtained after applying tests of the NIST suite on the two binary sequences produced by f_T in conjunction with the proposed discretization method, are presented.

Table 3. NIST results for bitstream_A

#	statistical test	passing ratio	p value	Obs.
1	Frequency	0.900555	0.993	S
2	Block frequency	0.980010	0.994	S
3	Cumulative sums (fwd)	0.874548	0.987	S
4	Cumulative sums (rev)	0.816537	0.988	S
5	Runs	0.137282	0.990	S
6	Longest run	0.935716	0.992	S
7	Rank	0.366918	0.985	S
8	FFT	0.964295	0.984	S
9	Non-overlapping template	0.401199	0.991	S
10	Overlapping template	0.171867	0.991	S
11	Universal	0.102526	0.988	S
12	Approximate entropy	0.334538	0.994	S
13	Random excursions	0.922036	0.985	S
14	Random excursions variant	0.551026	0.985	S
15	Serial (1)	0.224821	0.987	S
16	Serial (2)	0.262249	0.991	S
17	Linear complexity	0.202268	0.990	S

Table 4. NIST results for bitstream_B

#	statistical test	passing ratio	p value	Obs.
1	Frequency	0.837274	0.991	S
2	Block frequency	0.972253	0.990	S
3	Cumulative sums (fwd)	0.778587	0.985	S
4	Cumulative sums (rev)	0.782463	0.987	S
5	Runs	0.101988	0.991	S
6	Longest run	0.693142	0.990	S
7	Rank	0.433590	0.993	S
8	FFT	0.954154	0.987	S
9	Non-overlapping template	0.440048	0.984	S
10	Overlapping template	0.105305	0.994	S
11	Universal	0.588307	0.988	S
12	Approximate entropy	0.277585	0.992	S
13	Random excursions	0.629501	0.986	S
14	Random excursions variant	0.495347	0.985	S
15	Serial (1)	0.265567	0.990	S
16	Serial (2)	0.244259	0.991	S
17	Linear complexity	0.326749	0.986	S

Analyzing the results summarized in the two remembered tables it can be concluded that p-values for each statistical test are greater than 0,0001 and apear uniformly distributed in the interval [0, 1). Adding the fact that computed proportions for each test lies inside the confidence interval, bitstreams generated with the proposed method have very good cryptographic properties.

4.4. Complete statistical testing

Furthermore, considering all possible codings (6) of the intervals defined by the four thresholds (7) and using the rule (8), for the interpretation of each lexicographically generated permutation, numerical experiments (i.e. CrypTool, VRA, respectively the NIST standard battery tests) were performed on all possible permunations (9).

Each bitstreams pair, i.e. corresponding to each lexicographic permutation, subjected to same testing methodologies provides similar results, making the proposed, multi-level, discretization method suitable for cryptograohic applications, regardless on the encoding of thresholds.

$$(6) \qquad C(i) \in \{00, 01, 10, 11\} \; / \; \forall\, i = \overline{1, 4}$$

$$(7) \qquad \begin{array}{ll} a, & f_T \in [0, 0.25] \\ b, & f_T \in [0.25, 0.50] \\ c, & f_T \in [0.50, 0.75] \\ d, & f_T \in [0.75, 1] \end{array}$$

$$(8) \qquad \begin{array}{l} \textit{being given}: \\ C(1) \leftarrow 00,\; C(2) \leftarrow 01,\; C(3) \leftarrow 10,\; C(4) \leftarrow 11 \\ \textit{for any particular permutation of the set}: \\ \{C(1), C(2), C(3), C(4)\} \\ \textit{we can write the association matrix}: \\ \sigma \in S_4 = \begin{pmatrix} \sigma(C(1)) & \sigma(C(2)) & \sigma(C(3)) & \sigma(C(4)) \\ a & b & c & d \end{pmatrix} \end{array}$$

understanding that, for instance, 12th permutation can be written as:

$$\begin{aligned} \sigma &= \begin{pmatrix} C(4) & C(1) & C(3) & C(2) \\ a & b & c & d \end{pmatrix} \\ &= \begin{pmatrix} C(4) & C(1) & C(3) & C(2) \\ [0, 0.25] & [0.25, 0.50] & [0.50, 0.75] & [0.75, 1] \end{pmatrix} \end{aligned}$$

i.e. σ_{12} → [0, 0.25] is coded as 11, [0.25, 0.50] as 00, [0.50, 0.75] as 10 and [0.75, 1] as 01.

$$\begin{array}{l} \{a, b, c, d\}, \{a, b, d, c\}, \{a, c, b, d\}, \{a, c, d, b\} \\ \{a, d, b, c\}, \{a, d, c, b\}, \{b, a, c, d\}, \{b, a, d, c\} \\ \{b, c, a, d\}, \{b, c, d, a\}, \{b, d, a, c\}, \{b, d, c, a\} \\ \{c, a, b, d\}, \{c, a, d, b\}, \{c, b, a, d\}, \{c, b, d, a\} \\ \{c, d, a, b\}, \{c, d, b, a\}, \{d, a, b, c\}, \{d, a, c, b\} \\ \{d, b, a, c\}, \{d, b, c, a\}, \{d, c, a, b\}, \{d, c, b, a\} \end{array} \quad (9)$$

CONCLUSIONS

The desiderate was to achieve a higher / multiple flow of bits, much faster (i.e. speaking in terms of computational time), in order to use them for secure cryptographic application and/or other specific tasks such as generation of data streams for the study of traffic within WSNs, ATM or Telecom networks and stochastic computing.

In these sense, a new computational method for the transformation of real number chaotic sequences into multiple bitstreams, was proposed. For testing its viability for remembered applications one of the most popular maps was used: the tent map.

Wishing to preserve statistical properties of the newly formed multiple bitstreams, more than one exhaustive testing process of their randomness was performed, using specific/borrowed statistic suites. The results of statistical testing are encouraging and show that the proposed discretization method can be used for the development of secure cryptographic applications and other specific tasks.

As future work, actual usage of newly generated bitstreams in stochastic computing or in the study of traffic within WSNs and testing of the discretization method over others dynamical systems (e.g. cubic map, coupled chaotic maps etc.) is in the author’s interest.

REFERENCES

Alvarez, G., Montoya, F., Romera, M. and Pastor, G. (2012). Cryptanalyzing a discrete-time chaos synchronization secure communication system, *Chaos, Solutions and Fractals*, **Volume No. 21**, Issue No. 3, pp. 689÷694.

Alvarez, G. and Li, S. (2006). Some Basic Cryptographic Requirements for Chaos Based Cryptosystem, *Int. Journal of Bifurcation and Chaos*, **Volume No. 16**, pp. 2129÷2151.

Alvarez, G., Montoya, F., Romera, M. and Pastor, G. (2003). Cryptanalysis of an ergotic chaotic cipher, *Physics Letters A*, **Volume No. 311**, Issues No. 2÷3, pp. 172÷179.

Arroyo, D., Alvarez, G., Li, S., Li, C. and Nunez, J. (2008). Cryptanalysis of a discrete – time synchronous chaotic encryption system, *Physics Letters A*, **Volume No. 372**, Issue No. 7, pp. 1034÷1039.

Banerjee, S. and Pedersen, T. (2003), Design, Implementation and Use of the n-gram Statistics Package, *Proceeding of the 4th International Conference on Computational Linguistics and Intelligent text processing*, (CICLing), Mexico, 16÷22 February, pp. 370÷381.

DeCarlo, L.T. (1997). On the Meaning and Use of Kurtosis, *Psychological Methods*, **Volume No. 2**, Issue No. 3, pp. 292÷307.

Doane, D.P. and Seward, L.E. (2011). Measuring Skewness: A Forgotten Statistic?, *Journal of Statistics Education*, **Volume No. 19**, Issue No. 2, pp. 45÷63.

Jing, D., *et al*. (2005). Countermeasures Against Traffic Analysis Attacks in WSNs, *Proceedings of the 1st International Conference on Security and Privacy for Emerging Areas in Communications Networks*, Athens, Greece, 5÷9 September, pp. 113÷126.

Jorge, B.F. and Dulce, C. (2002). Recurrence Plots in Nonlinear Time Series Analysis: Free Software, *Journal of Statistical Software*, **Volume No. 7**, Issue No. 9, pp. 1÷18.

Kim, S.J., Umeno, K. and Hasegawa, A. (2004). *Corrections of the NIST Statistical Test Suite for Randomness”*, IACR Eprint archive, last accessed on [01.10.2012], http://eprint.iacr.org/2004/018.

Kocarev, L. and Lian, S. (2011). *Chaos - based Cryptography: Theory, Algorithms and Applications”*, Springer Verlag, pp. 227÷297.

Kocarev, L., Jakimoski, G., Stojanovski, T. and Parlitz, U. (1998). From Chaotic Maps to Encryption Schemes, *IEEE Proceeding of International Symposium Circuits and Systems*, (ISCAS), **Volume No. 4**, pp. 514÷517.

Li, S., Chen, G. and Mou, X. (2005). On the dynamical degradation of digital picewise linear chaotic maps, *International Journal of Bifurcation and Chaos*, **Volume No. 15**, Issue No. 10, pp. 3119÷3151.

Li, S., Mou, X., Cai, Y., Ji., Z. and Zhang, J. (2003). On the security of a chaotic encryption scheme: problems with computerized chaos in finite computing precision, *Computer Phys. Comm.*, **Volume No. 153**, Issue No. 1, pp. 52÷58.

Li, S., Mou, X. and Cai, Y. (2001). Pseudo-random Bit Generator Based on Couple Chaotic Systems and its Application in stream-ciphers cryptography, *Lecture Notes in Computer Science*, **Volume No. 2247**, pp. 316÷329.

Luca, A., Vlad, A., Badea, B. and Frunzete, M. (2009). A study on statistical independence in the tent map, *IEEE Proceeding of International Symposium on Signals, Circuits and Systems*, (ISSCS), Iaşi, România, pp. 1÷4.

Marwan, N., Romano, M.C., Thiel, M. and Kurths, J. (2007). Recurrence Plots for Analysis of Complex Systems, *Physics Reports*, **Volume No. 438**, Issues No. 5÷6, pp. 237÷329.

Muise, C., McIlraith, S., Baier, J.A. and Reimer, M. (2009). Exploiting n-gram Analysis to Predict Operator Sequences, *Proceeding of the 19th International Conference on Automated Planning and Scheduling*, (ICAPS), Thessaloniki, Greece, 19÷23 September, pp. 374÷377.

Patidar, V. and Sud, K.K. (2009). A novel pseudo random bit generator based on chaotic logistic map and its statistical testing, *Electronic Journal of Theoretical Physics*, (EJTP), **Volume No. 6**, Issue No. 20, pp. 327÷344.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., *et al.* (2010). *A statistical test suite for the validation of random number generators and pseudo random number generators for cryptographic applications*, NIST, SP 800-22, Revision 1a, April 2010, last accessed on [01.10.2012]: www.csrc.nist.gov.

Sebesta, V. (2007). Speeding Sequences Generated Using Asymmetrical Integer - Number Maps, *Journal of Radio engineering*, **Volume No. 16**, Issue No. 3, pp. 108÷112.

Shanon, C.E (1948). A mathematical theory of communication, *Bell Systems Technology Journal*, **Volume No. 27**, pp. 379÷423 & 623÷656.

Şerbănescu, A. and Râncu, C.I. (2008). *Systemes et signeaux face au chaos. Applications aux communications*, Military Technical Academy Publishing House, Bucharest, România.

*** *Cryptology with CrypTool*, last accessed on [01.10.2012], www.cryptool.com.

ABOUT THE AUTOR

Adrian - Viorel **DIACONU** is with Lumina - The University of South - East Europe, IT&C Dept., Bucharest, Romania, as Assistant Professor.

As PhD(c) at University Politehnica of Bucharest, but not only, Adrian - Viorel DIACONU has research interests focused mainly on designing of embedded systems and apps for WSNs, Robotics and Autonomous Vehicles.

Adrian - Viorel DIACONU has been with UESEL since February 2010, among his key responsibilities being found: INFOMATRIX's National Selections Coordinator, Jury Member in several national and international contests (e.g. firSTep, Infomatrix), establishing and maintaining various partnerships with other institutions of higher education (e.g. HIT, Netherlands) and local economic, industrial and social players (e.g. ANOSR, HP, FESTO) and more.