REVIEW

on PhD thesis for the educational and scientific degree “doctor” (PhD)

Topic: „Designing Boolean functions and digital sequences for cryptology and communications“

Author: Miroslav Marinov Dimitrov,
Section Mathematical Foundations of Informatics (MFI), Institute of mathematics and informatics (IMI), Bulgarian academy of sciences (BAS)

Scientific advisor: prof. Tsonka Stefanova Baicheva, DSc

Area of higher education: 4. Natural sciences, mathematics and informatics

Professional Field: 4.6. Informatics and computer sciences

By: prof. Nikolay Ivanov Yankov, DSc, Shumen University "Bishop Konstantin of Preslav”

I was appointed by order № 521/02.12.2022 of IMI’s director to be a member of this scientific jury and on the first session I was voted to write this review. I confirm that I have received all materials for this procedure according to LDASRB (the Law for the development of the academic staff in the Republic of Bulgaria). I do not have information for violation of the procedure, nor I’m aware of plagiarism in the presented PhD thesis.

1. PhD student “curriculum vitae”

Miroslav Dimitrov has a diploma issued in 2009 for bachelor of Informatics from Sofia University „St. Kl. Ohridski“. In 2016 he defended his Master’s thesis in the same university in the MS program „Information security in computer systems and networks“. In the presented very brief CV, unfortunately there is no information for his participation in scientific projects. The list of talks in scientific forums shows two items: at OCRT'17 in Sofia and CRYPTACUS in Portugal from 2018.

2. PhD Data

The doctoral studies started on 05.07.2017 for a 4-year period and concluded with the right to defend thesis according to decision by IMI’s scientific board (Protocol № 619/30.06.2022). The preliminary discussion of the dissertation, which I attended, took place on 11.11.2022 at an extended meeting of the MFI at the national coding seminar “Prof. S. Dodunekov”. By order of the Director of IMI, a scientific jury and the date of the defense have been determined. I believe that the procedure is regular and there are no violations.

3. Thesis and abstract data

The PhD thesis written in English has the following structure: introduction (6 pp.), main text (116 pp.) in 5 chapters divided in sections. There is a list of candidate's contributions, a list of 152 references, a list of candidate's publications, as well as lists for diagrams, tables and algorithms. The thesis also includes two appendices. The thesis meets the requirements of LDASRB and RALDASRB (Rules on the application of the Law for the development of the academic staff in the Republic of Bulgaria), as well as of Regulations on the terms and conditions for acquiring scientific degrees and for holding academic positions in BAS. The abstract (a total of 59 pp.) adequately reflects the main ideas and significant final results, which are described in the dissertation.

Although LDASRB does not require citation for doctoral defense, I consider excellent the fact of the 45 noted citations attached to the dissertation, of which more than half: 25 are from the last year. This reveals the great potential of the presented dissertation, as well as the major recognition by the international community of the significant scientific contributions of this work.

4. Review of the thesis

Chapter 2 of this PhD thesis is devoted to different strategies for constructing S-boxes: T_1 – pseudorandom generation; T_2 – algebraic or other deterministic approaches; T_3 – heuristics search method for finding a solution optimizing a pseudorandom box; and T_4 – a hybrid approach. A new algorithm is found based on stochastic hill climbing method with a proposed stopping point at reaching $\frac{n(n-1)}{4}$ cycles. Using this algorithm, a new maximal average coordinate nonlinearity value (ACNV) of 114,5 is achieved. Thus, it is shown (as was expected) that a structure (in this case using heuristics methods starting at pseudorandom S-boxes) gives better results than that achieved using chaotic functions. This is hardly a surprise to most researchers, as it's usually shown by practice, that random or chaotic structures are seldom optimal. In paragraph 2.5 is proven that the problem for finding optimal in the nonlinearity characteristics S-boxes can be transformed to standard binary optimization problem. Thus, it can be easily solved by standard algorithms implemented in most computer algebra systems. The translation of the problem is achieved by flipping two bits in the lookup table of the S-box. Some new terms are introduced: couplings, coordinate decomposition, degree of descendibility, among other. These new technics are later used to prove the connection of this problem to the solvable binary integer linear programming (BLIP) problem. In this case we do not need

to find an optimal bit-solution – just one solution suffices. As an application, an analysis on two well-known S-boxes Skipjack (USA) S_k and Kuznyechik K_k (used in a standardized by the Russian Federation symmetric cryptographic algorithm) is performed. Using the newly-proposed algorithm the nonlinearity of both boxes is improved from 100 to 102. For S_k the improvement is achieved by just 4 bit-flips, and for K_k the number of flips is (out of 2048 bits contained in both S-boxes). Starting the optimization process from the 8-by-8 S-box having the highest known average coordinate nonlinearity value (ACNV) of 114.5, a new S-box is found with ACNV of 116 and with overall nonlinearity 92.

The third chapter of the reviewed thesis is more applied. Based on the works by Aleksei Udovenko and Leo Perrin (and his doctoral thesis from 2017) and improvement is suggested. By using spectral analysis of S-boxes with criteria linear approximation, difference distribution, and auto-correlation tables an image can be rendered. This image can show (by coloring in red only these elements that meets the chosen criterion) to the observer different anomalies and symmetries that can be useful in reverse-engineering the S-box being analyzed. By choosing the appropriate criteria and carefully selecting the threshold values inexplicable structures are shown in many of the well-known S-boxes. A way to automize this process of finding anomalies is proposed, for example by using other S-box characteristics as: difference distribution table (DDT), auto-correlation tables (ACT) and XOR таблица (XORT).

Chapter 4's main subject of study are binary sequences. A novel shotgun hill climbing algorithm for finding sequences with optimal peak sidelobe levels (PSL) is proposed. This algorithm improves the computational and memory complexity to $O(n)$ levels. With its help record PSL breaking longer *m*-sequences are found. An application of the said algorithm is performed to rotated Legendre sequences. By achieving reduction in memory complexity of the PSL improving algorithm, as well as showing its high parallelizability, a CUDA implementation is proposed with the added advantage of simultaneously probing all of the sequence rotations. Thus, a gregarious improvement in the speed as well as minimization of the memory (compared to the standard algorithms, for example the one included in NymPi) is shown. All *m*-sequences with lengths $2^n - 1, n \in \{18, 19, 20\}$ are successfully exhausted. As a result of this improvements, a massive improvement in the problem of finding all optimal Legendre sequence up to given length is achieved. This problem is solved for lengths ≤ 432100.

The last chapter 5 of the dissertation is devoted to what is considered one of the most difficult optimization tasks – finding a method for improving the merit factor (MF) of binary sequences with given length. By using flip operation, an improvement of the algorithm for MF optimization is shown: Starting with an odd-length skew-symmetric sequence L, by flipping positions q and $n-q-1$ for given q, a new sequence L^q is constructed with the skew-symmetry being invariant. A set of 6 properties of the sequence L^q (Theorem 5.1.1) are discovered. These properties are helpful in transforming the formula into simplified form. The new proposed version of the algorithm achieves a memory reduction from $O(n^2)$ to $O(n)$ and at the same time the time complexity is not changed. Some new skew-symmetric binary sequences with $MF > 5$ are found. A new class of binary pseudo skew-symmetric sequences with large lengths (up to 10^5+1) is presented.

5. Scientific and applied-scientific contributions

The main scientific contributions are as follows:

- New theoretical techniques are proposed: coupling, coordinate decomposition, degree of descendibility and CELAT and they were used to solve the S-box nonlinearity optimization problem by translating to SAT problem which is solved using SAT solver.
- A new technique for S-box anomalies visualization is presented. This shows that a lot of the famous S-boxes have hidden structure and thus are unsafe.
- An algorithm for merit factor optimization is proposed. It reduces the memory requirement by an order of magnitude without affecting the calculation speed.
- A new algorithm based on heuristics shotgun hill climbing makes it possible to find binary sequences with small PSL values.
- A new class of binary sequences called skew-symmetric with even length is defined.

The applied-scientific contributions are:

- New bijective 8x8 S-boxes are constructed, possessing the maximal nonlinearity value of 116 on all 8 coordinates.
- Using the algorithm for pseudo-random generation of skew-symmetric sequences it is shown that there exist such with MF greater than 5 (for lengths up to 10^5+1). This rejects the Bernasconi's hypothesis stating that

a stochastic search cannot produce a sequence with length greater than 200 and MF greater than 5.

- New sequences with lengths between 106 and 300 having record breaking PSL values are constructed.
- A parallel algorithm is implemented and used to find PSL optimal binary sequences among all sequences and their rotations. This algorithm is used to find all optimal *m*-sequences with lengths $2^n - 1, n \in \{18, 19, 20\}$.
- All PSL-optimal Legendre sequences for lengths up to 432100 are found and a conjecture is stated that the number of all such sequences with lengths $N > 235723$ is sharp lower-bounded by $\sqrt{N}$.

I confirm that the scientific and applied contributions indicated in the dissertation are actually achieved.

6. Publications and participation in scientific forums

There are 9 papers in the list of publications and the vast majority of them are published in renown scientific journals included in the first quartile (Q1) of "Web of Science" and Scopus. The half of the published works having impact-factor have Miroslav Dimitrov as sole author and the other 3 are coauthored with two other researchers: the thesis advisor Prof. Tsonka Baicheva and Nikolay Nikolov, PhD.

Two of the papers are not yet published and are presented in Arxiv and the last paper is included in a conference proceedings volume. Exactly three of the published works are in IEEE Access – a journal included in top 85% (Scopus) in "computer science" that has IF of 3.476. Two papers are published in IEEE Signal Processing Letters having IF 3.109 and occupying top 90% in the applied mathematics category. The last peer-reviewed published paper is in IEEE Communications Letters with IF 3.436 again in the computer science realm. I believe that the participation of the doctoral student in all joint publications is equal to the other co-authors. The number of published papers significantly exceeds the requirements of RALDASRB.

7. Opinion, remarks and recommendations

A PhD thesis – as was mentioned in the preliminary discussion of the dissertation – is not just a collection of papers. In that sense, I found it strange that the introductory parts of the main chapters do not contain references to Miroslav's papers that are the meat and bone of this thesis. Such references are found only at the end of each section in Chapter 4. Moreover, at more than a few instances the reader is referred to these publications but not to the same text at

the corresponding place within the dissertation. In my opinion, a PhD thesis should be presented as a wholesome, coherent work.

8. Conclusion

This PhD thesis fully meets the requirements established by the LDASRB, as well as the regulations of BAS and IMI, so I confidently suggest to the esteemed scientific jury to vote for the educational and scientific degree "Doctor" to BE ACQUIRED by Miroslav Marinov Dimitrov in **Professional Field:** 4.6. Informatics and computer sciences.

Jury member:
……..

/prof. Nikolay Yankov, DSc/

Date: 05.01.2023