R E P O R T

on a Thesis for awarding the degree "Doctor"

Scientific field: 4. Natural sciences, mathematics and informatics

Professional field: 4.6. Informatics and Computer Science

Title: "Designing Boolean Functions and Digital Sequences for Cryptology and Communications"

Author: Miroslav Marinov Dimitrov

Overview

The presented thesis deals with a subject which can stated as construction and analysys of boolean functions, S-boxes and digital sequences. Problems of this kind can be related to the area of cryptography and data protection. The main problems treated in the thesis are combinatorial in nature. Such questions have obtained immense popularity in the last few years due to the ever increasing need of data protection and more specifically the construction of new and reliable block ciphers.

The results in this thesis are related first and foremost to the subject of symmetric cryptography. Every block cipher has a non-linear module called an S-box (substitution box) which guarantees that it is not vulnerable against the known cryptographic attacks. This module has to be designed very carefully and should satisfy strong mathematical criterea. This is the main motivation behind the thesis.

The thesis is based on the reserch carried out by the author in the past few years.

State of the current research

My general impression is that the author is well acquainted with the state of the art and the most recent results in the research on the treated problems. A good deal of the investigations are on problems that are considered to be central in the field of symmetric cryptography, but also have a huge theoretical importance. The author demonstrates deep knowledge of his field of research and capacity to apply his knowledge to the solution of important problems.

Methods

In his investigations the author uses a wide spectrum of software and mathematical tools. The results in the thesis are obtained by using a complicated orginal programs written by the author. Part of the statements have mathematical proofs that use ideas from the discrete mathematics, the integer optimization and the combinatorics.

Brief description of the thesis

The presented thesis amounts 190 pages of text and consists of an introduction, five chapters, two appendices and a list of references including 152 items. In what follows, I shall give a short description of the topics covered in this dissertation.

Chapter 1 is introductory and contains a brief description of the most important definitions and theoretical facts needed for the thesis. It starts with a explanation of the importance of S-boxes and why it is essential for these modules to be stable against the various cryptoanalytic methods: linear, differential, algebraic cryptanalysis. The importance of several purely mathematical problems is stressed that are related to the solution of this problem, e.g. the construction of (binary) sequences of low out-of-phase autocorrelation. Various classical characteristics are considered, notably the peak sidelobe level and the merit factor. Several interestiung problems are formulated that stay open at the present moment and are a challenge to the researchers, e.g. to determine the optimal values of the peak sidelobe level of sequences of length 84, minimization of the merit factor etc. Here the author formulates the main goal of this thesis: to investigate strategies for the construction and analysis of boolean functions. Further the author gives a brief description of the material covered in the five chapters of the thesis.

The original contributions of this thesis are contained in chapters 2, 3, 4 and 5.

Chapter 2 contains some preliminary definitions and facts on boolean functions and S-boxes needed in the following chapters. The author introduces farious normal forms for boolean functions, the transformation of Walsh-Hadamard, vector boolean functions, and objects related to them. The author analyses a wide range of actual S-boxes and provides a classification of the strategies used to construct S-boxes. Then he considers the problem of the construction of an S-box as a problem of the integer optimization. He introduces a program which optimizes the nonlinearity of a given S-box in minimal number of changes. This property turns out to be useful because it gives the possibility to focus on the weak components of an S-box without spoiling the rest. It is demonstrated that the suggested algorithm is highly effective by its application on the S-boxes of two well-known symmetric block ciphers – Skipjack and Kuznechik. An S-box with average coordinate nonlinearity of 116 is also demonstarted. The authors reports that all the attempts to construct such an S-box with average coordinate nonlinearity of 118 have failed so far.

In Chapter 3 the author considers various strategies for detection of hidden patterns and anomalies in popular S-boxes. Using this one can not only prove that the boxes are chaos based, but also obtain some additional information on their internal structure. This in turn

gives the possibility of reverse engineering, i.e. full recovery of the hidden structure of the S-box.

Chapter 4 deals with the problem of the construction of sequences with small autocorrelation, or, more precisely, optimisation of the peak sidelobe level of sequences of given length. In section 4.1 the author presents an effective euristic algorithm based on the so-called shotgun hill-climbing. The algorithm starts to work from a random binary sequence of length n. As an input it obtains a target value for the peak sidelobe level which is denoted by G. When the algorith reaches a sequence with this target value it stops. Otherwise it constructs a sequence which is at (Hamming) distance 1 from the original one. If all neighbour sequences do not improve on the achieved peak sidelobe level so far a "quake" is performed which is a random jump and the precedure starts anew. The number of the admissible quakes is bounded by some constant. This algorithm is applied for the generation of sequences of lengths between 106 and 300. For more than half of these lengths the author obtains an improvement of the optimal values known so far. In the next section a method for the generation of binary sequences with a low value of the peak sidelobe level is sugested. The complexity of this algorithm is linear on space and time (from the length of the sequence). Experiments are made at which plenty of sequences with record-breaking PSL values are obtained. In the next section the author present various hybrid algorithms for the generation of binary sequences with low autocorrelation.

In Chapter 5 the author considers the problem of optimizing the so-called merit factor defined as the ratio of the energy of the mainlobe level to the energy of the sidelobe levels:

$$MF(B) = \frac{C_0(B)}{2\sum_{u=1}^{n-1} |C_u(B)|^2}$$

Furthermore, new classes of binary sequences of high merit factor are constructed. The author restricts his research on the class of so-called skew symmetric sequences introduced by Golay. Long sequences ($n > 200$) are constructed that exceed the merit factor predicted in Bernasconi's hypothesis. An algorithm for obtaining long binary sequences of high merit factor is also presented.

Main results

The main contributions of this thesis amount to the following:

(1) A detailed analysis of a large number of S-boxes is made and it is shown that a large number of chaos-based S-boxes are vulnerable to linear cryptanalysis.

(2) The exitence of 8×8 bijective S-boxes is proved that are of maximal nonlinearity 116 on all coordinates.

(3) Using S-box spectrography, a strategy for detection of hidden regularity patterns in S-boxes is suggested which can be used for reverse engineering.

(4) An effective euristic algorithm for the construction of binary sequences of low peak sidelobe level is suggested. Sequences of various lengths between 106 and 300 are constructed that have a record-breaking peak sidelobe level.

(5) New classes of binary sequences of high merit factor are constructed.

(6) A new class of finite binary sequences of even is constructed that has alternating autocorrelation values with absolute value equal to 1.

(7) An algorithm for obtaining long binary sequences of high merit factor is constructed.

Remarks and comments

I have the following remarks, questions and comments related to this thesis:

(1) In Definition 2.2.2 bijective S-boxes are possible only when $m = n$.

(2) In the formula for $MF(B)$ on page 48: why do we need the constant 2 in the denominator? On the other hand, what is the point of having C_0 in the same formula? It is always equal to the length of the sequence and hence a constant.

(3) What is the meaning of the merit factor. It is seems to me that it is enough to define the problem as one of minimizing the sum of the squares of C_u.

(4) In Definition 2.2.7: What do we call the length of a binary function?

(5) Definitions 2.2.7 and 2.2.8 are not very clear.

(6) It would be nice if the text contains a definition of the notion nonlinearity of a binary function, and not just non-linearity of an S-box.

Publications related to the thesis

The results in this thesis are published in nine papers, seven of which have already appeared in print and two are deposited in arXiv. The respective journals in which the papers are published are the following:

- IEEE Access – 3 papers;

- IEEE Signal Processing Letters – 2 papers;

- IEEE Communication Letters – 1 paper;

- Proceedings of the International Conference on Information Technologies (InfoTech) – 1 paper;

- arXiv – 2 papers.

Six of the papers are in journals with an impact factor which exceeds 3. The journals with impact factor are all in Q2. From the remaining papers one is in the proceedings of a scientific conference and two are deposited in arXiv.

In five of the papers the candidate is the sole author, in three of them he has two coauthors, and one is with three coauthors.

The scientific output of Miroslav Dimitrov is impressing. It includes papers published only in the last three years in very prestigeous journals. This speaks of a high level of the research conducted by the candidate. The presented publications do not just satisfy, but also exceed by far the minimal national criterea for awarding the educational and scientific degree "doctor".

Authorship of the obtained results

In five of the papers the candidate is the only author. In the remaining four he has coauthors. Since I have been following the scientific output of the author for a quite some time, I have no doubt that his contribution in the joint papers is significant.

Citations

The candidate has provided a list of 48 citations of the papers used in this thesis. Although the papers on which the thesis is based appeared shortly before the completion of the thesis they have already attracted the attention of the researchers in the field and are extensively cited. This shows the recognition they obtained by the professionals in the field. I have no doubt that the citations of the author's research will increase in the future.

Authors summary

The author's summary is made according to the existing regulations and reflects properly the main results and contributions of this thesis.

Conclusion

This thesis is focused on problems from symmetric cryptography and the theory of boolean functions and binary sequences. This work does not only answer open problems of principal importance, but also motivates new directions for an ongoing research. I am deeply convinced that the presented thesis **"Designing Boolean Functions and Digital Sequences for Cryptology and Communications"** by **Miroslav Marinov Dimitrov** contains results

that are an original contribution to the symmetric cryptography and the theory of binary sequences. The candidate demonstrates deep knowledge of the theory and capacity to develop it in new and important ways. With this, he meets the legal national requirements prescribed by the law, as well as the specific requirements of the Institute of Mathematics and Informatics of Bulgarian Academy of Sciences for the professional field 4.6 "Informatics and Computer Science". I assess **positively** the presented PhD Thesis and recommend to this panel to award **Miroslav Marinov Dimitrov** the scientific degree "Doctor" in the scientific field 4. Natural Sciences, Mathematics and Informatics, Professional field 4.6 "Informatics and Computer Science".

Sofia, 12.01.2023 Member of the Scientific Panel:

(Prof. DSc Ivan Landjev)