JET BRAINS RESEARCH

Труды лаборатории криптографии

2019-2020 учебный год

НОВОСИБИРСК 2020

ОГЛАВЛЕНИЕ

Состав лаборатории:

Токарева Наталья Николаевна
Заведующая лабораторией криптографии JetBrains Research, к.ф.-м.н., с.н.с. Института математики им. С.Л.Соболева СО РАН, доцент кафедры компьютерных систем ФИТ НГУ, кафедры теоретической кибернетики ММФ НГУ и кафедры дискретной математики и информатики СУНЦ НГУ.
E-mail: tokareva@math.nsc.ru

Коломеец Николай Александрович

Исследователь лаборатории криптографии JetBrains Research, к.ф.-м.н., н.с. Института математики им.С.Л.Соболева СО РАН, ассистент кафедры параллельных вычислений ФИТ НГУ и кафедры теоретической кибернетики ММФ НГУ, преподаватель кафедры дискретной математики и информатики СУНЦ НГУ.
E-mail: kolomeec@math.nsc.ru

Городилова Анастасия Александровна

Исследователь лаборатории криптографии JetBrains Research, к.ф.-м.н., н.с. Института математики им.С.Л.Соболева СО РАН, старший преподаватель кафедры теоретической кибернетики ММФ НГУ и кафедры дискретной математики и информатики СУНЦ НГУ.
E-mail: gorodilova@math.nsc.ru

Калгин Константин Викторович

Исследователь лаборатории криптографии JetBrains Research, к.ф.-м.н., н.с. Института вычислительной математики и математической геофизики СО РАН, старший преподаватель кафедры параллельных вычислений ФИТ НГУ.
E-mail: kalginkv@gmail.com

Идрисова Валерия Александровна

Исследователь лаборатории криптографии JetBrains Research, к.ф.-м.н., н.с. Института математики им.С.Л.Соболева СО РАН, ассистент кафедры теоретической кибернетики ММФ НГУ.
E-mail: vvitkup@yandex.ru

Куценко Александр Владимирович

Исследователь лаборатории криптографии JetBrains Research, аспирант ММФ НГУ, ассистент кафедры теоретической кибернетики ММФ НГУ, преподаватель кафедры дискретной математики и информатики СУНЦ НГУ.
E-mail: alexandrkutsenko@bk.ru

Облаухов Алексей Константинович

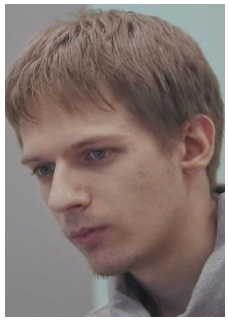
Исследователь лаборатории криптографии JetBrains Research, аспирант Института математики им. С.Л.Соболева СО РАН, н.с. Института математики им.С.Л.Соболева СО РАН, ассистент кафедры теоретической кибернетики ММФ НГУ, преподаватель кафедры дискретной математики и информатики СУНЦ НГУ.
E-mail: nskuber94@gmail.com

Сазонова Полина Андреевна

Исследователь лаборатории криптографии JetBrains Research, аспирант ФИТ НГУ, ассистент кафедры компьютерных систем ФИТ НГУ, преподаватель ЭФ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.
E-mail: psazonova@gmail.com

Кондырев Дмитрий Олегович

Исследователь лаборатории криптографии JetBrains Research, аспирант ФИТ НГУ, преподаватель ФИТ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.
E-mail: dkondyrev@gmail.com



Ткачев Александр Витальевич

Исследователь лаборатории криптографии JetBrains Research, аспирант ФИТ НГУ, преподаватель ФИТ НГУ.
E-mail: alexander@tkachov.ru

Косточка Светлана Владимировна

Исследователь лаборатории криптографии JetBrains Research, м.н.с. Института математики им.С.Л.Соболева СО РАН, тренер ММФ и ФИТ НГУ.

Бадер Дмитрий Александрович

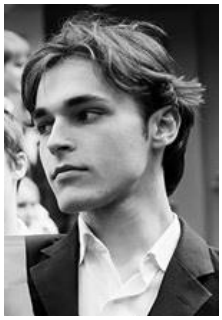
Исследователь лаборатории криптографии JetBrains, магистрант 2-го курса ММФ НГУ.

Белоусова Алина Александровна

Исследователь лаборатории криптографии JetBrains Research, магистрантка 2-го курса ММФ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.

Бонич Татьяна Андреевна

Исследователь лаборатории криптографии JetBrains, магистрантка 1-го курса ММФ НГУ.



Валиахметов Илья Вадимович

Исследователь лаборатории криптографии JetBrains, магистрант 1-го курса ФИТ НГУ.

Доронин Артемий Евгеньевич

Исследователь лаборатории криптографии JetBrains Research, магистрант 2-го курса ММФ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.

Журавлев Вячеслав Александрович

Исследователь лаборатории криптографии JetBrains Research, магистрант 1-го курса ММФ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.

Завалишина Елена Владимировна

Исследователь лаборатории криптографии JetBrains Research, магистрантка 1-го курса ФИТ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН, преподаватель кафедры дискретной математики и информатики СУНЦ НГУ.

Максимлюк Юлия Павловна

Исследователь лаборатории криптографии JetBrains Research, магистрантка 2-го курса ММФ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.

Панферов Матвей Андреевич

Исследователь лаборатории криптографии JetBrains, магистрант 1-го курса ММФ НГУ.

Пинтус Григорий Михайлович

Исследователь лаборатории криптографии JetBrains, магистрант 2-го курса ММФ НГУ.

Шапоренко Александр Сергеевич

Исследователь лаборатории криптографии JetBrains Research, магистрант 2-го курса ММФ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.

Атутова Наталья Дмитриевна

Исследователь лаборатории криптографии JetBrains, студентка 2-го курса ММФ НГУ.

Большим Максим Анатольевич

Исследователь лаборатории криптографии JetBrains, студент 2-го курса ФИТ НГУ.

Жантуликов Булат Фаритович

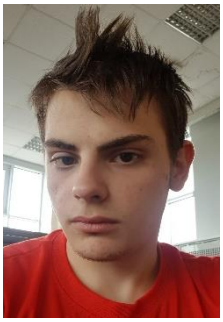
Исследователь лаборатории криптографии JetBrains Research, студент 1-го курса ФИТ НГУ, преподаватель кафедры дискретной математики и информатики СУНЦ НГУ.

Запольский Максим Михайлович

Исследователь лаборатории криптографии JetBrains, студент 3-го курса ММФ НГУ.

Зюбина Дарья Александровна

Исследователь лаборатории криптографии JetBrains Research, студентка 3-го курса ФИТ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.



Никифоров Владислав Сергеевич

Исследователь лаборатории криптографии JetBrains, студент 2-го курса ФИТ НГУ.

Парфёнов Денис Романович

Исследователь лаборатории криптографии JetBrains, студент 3-го курса ФИТ НГУ.

Софронова Дарья Алексеевна

Исследователь лаборатории криптографии JetBrains, студентка 2-го курса ФИТ НГУ.

Сутормин Иван Александрович

следователь лаборатории криптографии JetBrains Research, студент 4-го курса ММФ НГУ, м.н.с. Института математики им.С.Л.Соболева СО РАН.

Сычев Алексей Дмитриевич

Исследователь лаборатории криптографии JetBrains, студент 4-го курса ММФ НГУ.

Хильчук Ирина Сергеевна

Исследователь лаборатории криптографии JetBrains, студентка 4-го курса ММФ НГУ.

Ходзицкий Артем Федорович

Исследователь лаборатории криптографии JetBrains, студент 2-го курса ММФ НГУ.

Черников Василий Викторович

Исследователь лаборатории криптографии JetBrains, студент 2-го курса ФИТ НГУ.

NSUCRYPTO 2019

Наша команда выступает основным организатором международной олимпиады NSUCRYPTO.

NSUCRYPTO – единственная международная олимпиада по криптографии, которая объединяет как школьников и студентов, так и профессионалов. За время существования олимпиады (с 2014 года) в ней приняли участие более 1600 участников из 56 стран мира (среди них – страны ЕС, страны СНГ, Канада, Китай, Индия, ЮАР, Иран, Индонезия, Вьетнам и др.). По итогам каждой олимпиады публикуются научные статьи с разбором проблем, предложенных участникам, в том числе – нерешенных, требующих отдельного научного исследования. Отличительная черта олимпиады – включение в число ее задач нерешенных проблем криптографии и информационной безопасности, предложенных ведущими специалистами в данной области. Это как раз соответствует цели олимпиады – привлечь молодых исследователей к современным вопросам криптографии и помочь им сделать свой профессиональный выбор.

Олимпиада NSUCRYPTO – Non Stop University Crypto – проходит ежегодно, принять в ней участие может любой желающий. Официальный язык олимпиады – английский. Сайт – https://nsucrypto.nsu.ru.

Олимпиада зародилась в Новосибирском Академгородке. В 2019 году она проходила с 13 по 21 октября в два независимых этапа: личный и командный. Школьники Академгородка и студенты НГУ приняли в ней активное участие.

Призёры олимпиады NSUCRYPTO

Организации:

- Novosibirsk State University
- Sobolev Institute of Mathematics
- KU Leuven
- Belarusian State University
- Tomsk State University

Программный комитет:

- Gennadiy Agibalov (Tomsk State University, Russia)
- Sergey Agievich (Belarusian State University, Belarus')
- Lilya Budaghyan (University of Bergen, Norway)
- Anne Canteaut (INRIA Paris, France)
- Claude Carlet (University of Paris 8, France)
- Joan Daemen (Radboud University, The Netherlands)
- Sugata Gangopadhyay (Indian Institute of Technology Roorkee, India)
- Evgeny Gorkunov (Novosibirsk State University, Russia)
- Anastasiya Gorodilova (Sobolev Institute of Mathematics, Russia) co-chair
- Tor Helleseth (University of Bergen, Norway)
- Xiang-dong Hou (University of South Florida, USA)
- Valeriya Idrisova (Sobolev Institute of Mathematics, Russia)
- Nikolay Kolomeec (Sobolev Institute of Mathematics, Russia)
- Alexander Kutsenko (Novosibirsk State University, Russia)
- Roman Lebedev (Novosibirsk State University, Russia)
- Nicky Mouha (Computer Security Division of NIST, USA)
- Svetla Nikova (KU Leuven, Belgium)
- Alexey Oblaukhov (Sobolev Institute of Mathematics, Russia)
- Irina Pankratova (Tomsk State University, Russia)
- Stjepan Picek (Delft University of Technology, The Netherlands)
- Bart Preneel (KU Leuven, Belgium)
- Marina Pudovkina (Bauman Moscow State Technical University, Russia)
- Vincent Rijmen (KU Leuven, Belgium; University of Bergen, Norway)
- Razvan Rosie (University of Luxembourg, Luxembourg)
- Alexander A. Semenov (Institute for System Dynamics and Control Theory, Russia)
- Francesco Sica (Nazarbayev University, Kazakhstan)
- Pante Stanica (Naval Postgraduate School, USA)
- Natalia Tokareva (Novosibirsk State University, Russia)
- Meltem Turan (National Institute of Standards and Technology, USA)
- Aleksei Udovenko (CryptoExperts, France)

Председатель программного комитета:

- Natalia Tokareva (Novosibirsk State University, Russia)

ЗАЩИТЫ ВЫПУСКНЫХ РАБОТ СТУДЕНТОВ И АСПИРАНТОВ ЛАБОРАТОРИИ В 2020 ГОДУ:

Выпускная работа аспиранта:

- Сазонова Полина Андреевна (рук. - Токарева Н.Н.)
  Разработка методов для систематизации блокчейн-технологий

Магистерские диссертации:

- Бадер Дмитрий Александрович (рук. - Токарева Н.Н., Сазонова П.А.)
  Development of methods for transaction analysis in blockchain network - Разработка методов анализа транзакций в сети блокчейн

- Белоусова Алина Александровна (рук. - Токарева Н.Н.)
  Lai-Massey block ciphers and their properties - Блочные шифры типа Лая-Мэсси и их свойства

- Доронин Артемий Евгеньевич (рук. - Калгин К.В., Токарева Н.Н.)
  Construction of cryptographic Boolean functions using SAT-solvers - Поиск криптографических булевых функций с помощью SAT-решателей

- Максимлюк Юлия Павловна (рук. - Евдокимов А.А.)
  Исследование метрического дополнения упаковки цепей в булевом кубе - Study of the metric complements to chains packaging in Boolean cube

- Пинтус Георгий Михайлович (рук. - Токарева Н.Н., Куценко А.В.)
  On decomposition of vectorial Boolean functions for threshold implementation - О декомпозиции векторных булевых функций для пороговой реализации

- Шапоренко Александр Сергеевич (рук. - Токарева Н.Н., Куценко А.В.)
  Quaternary bent functions: properties and connection to Boolean bent functions - Кватернарные бент-функции: свойства и взаимосвязь с булевыми бент-функциями

Бакалаврские диссертации:

- Сутормин Иван Александрович (рук. - Коломеец Н.А., Городилова А.А.)
  Конструкции сбалансированных булевых функций с высокой нелинейностью

- Сычев Алексей Дмитриевич (рук. - Токарева Н.Н., Сазонова П.А.)
  Разработка протокола меж-блокчейн взаимодействия для случая залогового удержания

- Хильчук Ирина Сергеевна (рук. - Токарева Н.Н.)
  Построение и анализ S-блоков симметричных шифров

MASTER IN CRYPTO

В 2020 году состоялся выпуск студентов, проходивших обучение по уникальной магистерской программе "Master in Cryptography" на базе ММФ НГУ, полностью организованной нашей командой.

"Master in Crypto" - первая в России англоязычная магистратура по криптографии. Ее основная цель — привлечь сильных студентов со всего мира для глубокого изучения теоретических и практических аспектов современной криптографии и дальнейшего вовлечения перспективных студентов в научно-исследовательскую деятельность в данной области.

Для чтения лекций были приглашены российские и зарубежные специалисты в области криптографии.

Обучение проводилось на английском языке.

Курсы, включённые в программу:

- Algebra and finite fields: special aspects
- Discrete mathematics
- Theory of probability and mathematical statistics
- Numerical methods in cryptography
- Information theory and cryptography. Introduction
- Foundations of symmetric cryptography
- Cryptographic Boolean functions
- Cipher design
- Cryptanalysis of symmetric system
- Asymmetric cryptography and cryptanalysis
- Blockchain: math problem and applications
- Quantum and postquantum cryptography
- Practical applications of cryptography
- Historical and legal aspects of cryptography

Выпускники программы «Master in crypto».

Новости лаборатории в 2019-2020 учебном году:

- *НГУ запустил курс по криптографии на Coursera*

  Новый курс «Cryptography: Boolean functions and related problems» предназначен для тех, кто интересуется алгоритмами и методами шифрования, а также владеет английским языком и математическими знаниями. - Курсов по криптографии существует довольно много. В том числе вводных и базовых. Особенностью нашего курса является то, что, начав с простых основ, мы быстро перейдем к современным проблемам криптографии и поможем слушателям сделать первые шаги в научной криптографии, получить новые результаты. В двух словах цель нашего курса — слушателя-любителя криптографии превратить в исследователя. Эту цель мы преследуем и в проводимой нами ежегодно Международной олимпиаде по криптографии NSUCRYPTO, и в Летней школе по криптографии и информационной безопасности, — отметила доцент ФИТ и ММФ НГУ, руководитель новосибирского Криптографического центра Наталья Токарева. Курс состоит из пяти модулей, во время изучения которых слушатели познакомятся с самыми современными методами криптографии, научатся применять разные виды булевых функций для создания надежных шифров, а также узнают, какие проблемы могут при этом возникнуть и как с ними справиться. В число преподавателей вошли не только сотрудники НГУ, но и Степан Пичек, доцент, доктор наук из Делфтского университета, который занимает 50-ю строчку QS World University Rankings.

- *Открытые лекции Computer Science клуба при НГУ "Основы криптоанализа"*

Лекции читает Наталья Николаевна Токарева, к.ф.-м.н., с.н.с. Института математики им. С.Л.Соболева, руководитель Лаборатории криптографии JetBrains Research, руководитель Криптографического Центра (Новосибирск), доцент НГУ.

Курс в видео-формате доступен по ссылке:

https://nsk.compsciclub.ru/en/courses/cryptoanalysis/nsk/2020-spring/classes/

- *Сотрудники лаборатории - Александр Куценко и Алексей Облаухов - прошли научную стажировку в университете Бергена*

  С 3 февраля по 1 марта 2020 года Александр Куценко и Алексей Облаухов стажировались в научно-исследовательском Selmer Center in Secure Communication Университета Бергена (Норвегия). За это время были проведены совместные научные исследования, трижды ребята выступили на семинаре лаборатории:

  13.02.2020 - A. Kutsenko, «Self-dual bent functions: characterization and metrical properties». Рассмотрены известные свойства самодуальных бент-функций. Изложены полученные метрические свойства: минимальное расстояние Хэмминга между самодуальными бент-функциями, спектр расстояний Хэмминга между функциями из класса Мэйорана-МакФарланда. Доказана метрическая регулярность и найдено метрическое дополнение множества самодуальных бент-функций.

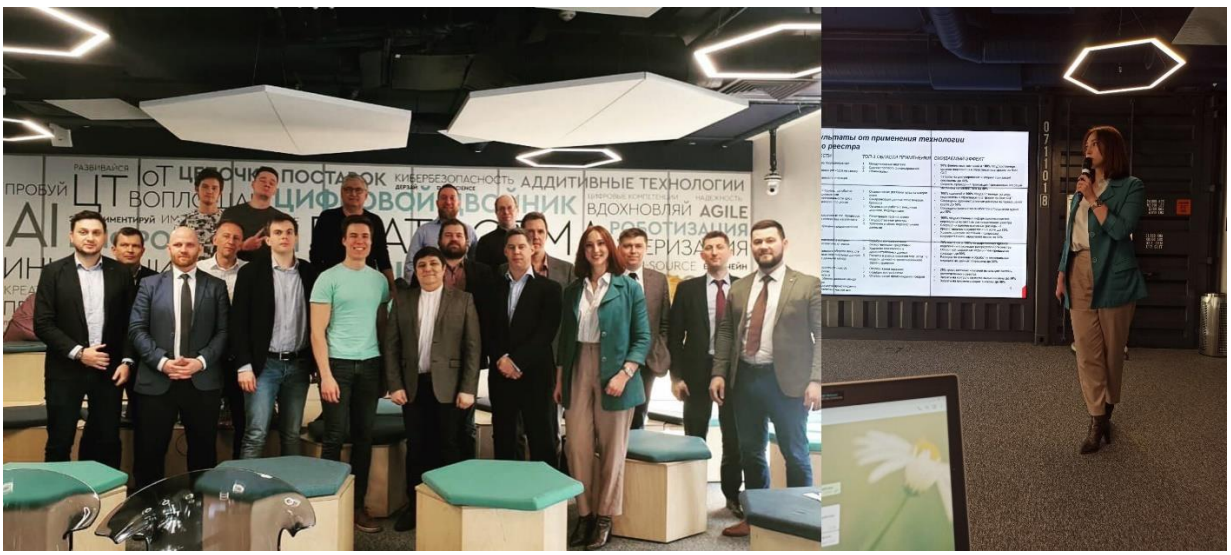
  20.02.2020 - A. Kutsenko, «The group of automorphisms of the set of self-dual bent functions». Приведены полученные результаты по изометрическим отображениям множества

самодуальных бент-функций. Доказано, что группы автоморфизмов множеств самодуальных и анти-самодуальных бент-функций совпадают. Полностью описана группа автоморфизмов множества самодуальных бент-функций.

27.02.2020 - A. Oblaukhov, «Metric complements and metric regularity in the Boolean cube». Излагаются полученные результаты, затрагивающие свойства метрических дополнений подмножеств булева куба. Найден общий вид метрического дополнения линейного подпространства булева куба. Получена нижняя оценка на мощность максимального метрически регулярного множества. Доказана метрическая регулярность кодов Рида-Маллера RM(k, m) для случая k>=m−3.

- *Сотрудница лаборатории Полина Сазонова выступила с приглашенным докладом на встрече "Digital Standup blockchain" в Доме Инноваций Газпром нефть*

Практическими приложениями блокчейн-технологий и научными результатами в данной области активно интересуются многие компании. На встрече 27 февраля, проходившей в Доме Инноваций Газпром нефть (г. Санкт-Петербург), Полина Сазонова выступила приглашенным докладчиком из Новосибирска. Она представила доклад «Технологии распределенного реестра как основа экономики России». На встрече обсуждались результаты в области блокчейн-технологий и вопросы их внедрения с участием представителей Почты России, Сбербанка, Сибинтека, Норникеля, Университета Иннополис и др. организаций. На круглом столе, организованном после докладов, участники обсудили перспективы развития блокчейн-технологии и существующие барьеры ее применения.



- МНСК-2020. В Международной студенческой конференции, ежегодно проводимой НГУ, в 2020 году приняли участие 12 студентов лаборатории.

  Поздравляем наших ребят, занявших призовые места на МНСК!

  В этом году она впервые проходила в режиме онлайн. Призёрами стали:

  Секция "Информ. технологии". Подсекция "Информационная безопасность"

  - Елена Завалишина (диплом I степени) "Криптоанализ базовой версии криптографической системы с открытым ключом, основанной на сложности решения системы полиномиальных уравнений в целых числах
  - Дарья Зюбина (диплом II степени) "Криптографические свойства S-блока, построенного на основе булевой функции и перестановки"
  - Дмитрий Бадер (диплом III степени) "Разработка методов анализа блокчейн сетей"

  Секция "Математика". Подсекция "Теоретическая кибернетика"

  - Александр Шапоренко (диплом III степени) "Связь кватернарных и булевых бент-функций"

- Студентам лаборатории криптографии присуждена Премии Ляпунова I степени в конкурсе дипломных работ 2020 года!

  - Ю.П.Максимлюк "Исследование метрического дополнения упаковки цепей в булевом кубе"(науч.рук.А.А.Евдокимов)

  - А.С.Шапоренко "Quaternary bent functions: properties and connection to Boolean bent functions" (науч.рук. А.В.Куценко, Н.Н.Токарева)

- Сотрудник лаборатории Александр Куценко принял участие в научно-технической конференции "Информационная безопасность" на базе Военного инновационного технополиса "Эра"

  19-20 марта Министерство обороны Российской Федерации и Военный инновационный технополис «ЭРА» провели вторую Всероссийскую научно-практическую конференцию «Состояние и перспективы развития науки по направлению «Информационная безопасность». Ее участниками стали представители военных и гражданских вузов, научно-исследовательских организаций, предприятий военно-промышленного комплекса и IT-компаний. Они представили свои наработки в области кибербезопасности.

Летняя школа-конференция «Криптография и информационная безопасность» 2020

Летняя школа-конференция «Криптография и информационная безопасность» — традиционное мероприятие, проходящее в стенах НГУ каждый год. Организаторами школы-конференции выступают Криптографический центр (Новосибирск), лаборатория криптографии JetBrains Research, Международный математический Центр в Академгородке, организаторы международной олимпиады NSUCRYPTO, Факультет информационных технологий и Механико-математический факультет. Основатель летних школ по криптографии — профессор ФИТ Сергей Федорович Кренделев.

Участие в школе-конференции принимали студенты, выпускники школ и школьники 11 классов. Школа проходила с 9 по 27 июля в дистанционном формате.

В течение трех недель со студентами работали около 15 преподавателей. Студенты принимали участие в лекциях, командной и индивидуальной работе в проектах, связанной с решением исследовательских задач в области криптографии и информационной безопасности, в спортивных занятиях. Одно из важнейших событий школы-конференции – круглый стол по современным проблемам криптографии. Темы проектов были связаны с различными вопросами современной криптографии и информационной безопасности: от разработки современных методов криптоанализа, построения шифров, квантовой криптографии до создания систем аналитической разведки с открытым кодом. В 2020 году в рамках летней школы-конференции со студентами работали преподаватели из России, Европы и США, в том числе авторы международных стандартов в области криптографии. Часть школы-конференции проходила на английском языке.

Школу успешно закончили 52 студента (ровно вдвое больше, чем в прошлом году). Это студенты из НГУ (ФИТ, ММФ), ТГУ, ТюмГУ, НГТУ, МФТИ, СибГУ (Красноярск), Алтайского ГТУ, Ереванского госуниверситета (Армения) и пять школьников: из Санкт-Петербурга, Бердска, Луховиц, Стерлитамака (респ. Башкортостан), Усть-Каменогорска (Казахстан). Всего на школу было подано 96 заявок.

Руководитель школы — к.ф.-м.н. Токарева Наталья Николаевна, доцент кафедры компьютерных систем ФИТ, кафедры теоретической кибернетики ММФ, с.н.с. ИМ СО РАН, зав. лаб. криптографии JetBrains Research.

Летняя школа-конференция "Криптография и информационная безопасность" 2020 www.crypto.nsu.ru

Лекторы и преподаватели школы:

- Nicky Mouha (USA) - PhD, научный сотрудник отдела компьютерной безопасности Национального института стандартов и технологий США (NIST);
- Агиевич Сергей Валерьевич (республика Беларусь) - к.ф.-м.н., заведующий НИЛ проблем безопасности информационных технологий НИИ прикладных проблем математики и информатики Белорусского государственного университета (г. Минск, Беларусь);
- Городилова Анастасия Александровна - к.ф.-м.н., старший преподаватель кафедры теоретической кибернетики ММФ НГУ, н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;
- Калгин Константин Викторович - к.ф.-м.н., старший преподаватель кафедры параллельного программирования ФИТ НГУ, м.н.с. ИВМиМГ, н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;
- Колегов Денис Николаевич - к.т.н., доцент кафедры компьютерной безопасности ТГУ, главный разработчик облачной платформы кибербезопасности компании Bi.Zone (Томск);
- Коломеец Николай Александрович - к.ф.-м.н., ассистент кафедры теоретической кибернетики ММФ НГУ, кафедры параллельного программирования ФИТ НГУ, н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;
- Кондырев Дмитрий Олегович - аспирант ФИТ НГУ, ассистент кафедры систем информатики ФИТ НГУ, м.н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;
- Куценко Александр Владимирович - аспирант ММФ НГУ, ассистент кафедры теоретической кибернетики ММФ НГУ, м.н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;
- Малыгина Екатерина Сергеевна - к.ф.-м.н., доцент Балтийского федерального университета им. Иммануила Канта (Калининград);
- Николаев Антон Анатольевич - студент кафедры компьютерной безопасности ТГУ, разработчик сервисов анализа защищенности Bi.Zone, главный разработчик фрэймворка Grinder (Томск);
- Облаухов Алексей Константинович - аспирант ИМ СО РАН, ассистент кафедры теоретической кибернетики ММФ НГУ, м.н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;
- Пудовкина Марина Александровна - д.ф.-м.н., профессор МГТУ им. Баумана (Москва);
- Сазонова Полина Андреевна - аспирантка ФИТ НГУ, ассистент кафедры общей информатики ФИТ НГУ, м.н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;
- Токарева Наталья Николаевна - к.ф.-м.н., доцент кафедры компьютерных систем ФИТ, кафедры теоретической кибернетики ММФ, с.н.с. ИМ СО РАН, зав. лаб. криптографии JetBrains Research.
- Завалишина Елена Владимировна - магистрантка ФИТ НГУ, м.н.с. ИМ СО РАН, сотрудник лаборатории криптографии JetBrains Research;

Лекции школы:

1. Криптография: быстрый старт - Токарева Н.Н.
2. Постквантовая криптография: вызов брошен? - Куценко А.В.
3. Nonstop University CRYPTO: как совместить игру и науку в формате олимпиады? - Городилова А.А.
4. Криптография и криптоанализ с открытым ключом - Токарева Н.Н.
5. Практические аспекты компьютерной безопасности - Колегов Д.Н.
6. On proving security against differential cryptanalysis 1 - Nicky Mouha
7. On proving security against differential cryptanalysis 2 - Nicky Mouha
8. Основы алгебраического криптоанализа - Куценко А.В.
9. Основы технологии блокчейн - Сазонова П.А.
10. ARX-based cryptography 1 - Nicky Mouha
11. ARX-based cryptography 2 - Nicky Mouha
12. Сетевое сканирование и анализ угроз - Колегов Д.Н., Николаев А.А.
13. Методы и техники анализа веб-приложений - Колегов Д.Н., Николаев А.А.
14. Блокчейн изнутри - Кондырев Д.О
15. SAT-решатели и их приложения в криптографии - Калгин К.В.
16. Безопасность систем машинного обучения - Колегов Д.Н., Николаев А.А.
17. Введение в поиск пропавших людей с использованием методов и техник разведки на основе открытых источников - Колегов Д.Н., Николаев А.А.
18. Хэш-функции: построение и анализ - Коломеец Н.А.
19. Алгебраический анализ LRX-шифров - Куценко А.В.
20. Методы машинного обучения в системах разведки на основе открытых источников - Колегов Д.Н., Николаев А.А.
21. Разработка инструментов с применением методов разведки на основе открытых источников - Колегов Д.Н., Николаев А.А.
22. Квантовый криптоанализ: первое приближение - Куценко А.В.
23. Как стать специалистом по компьютерной безопасности - Колегов Д.Н.
24. Введение в теорию эллиптических кривых - Малыгина Е.С.
25. Введение в криптографию на эллиптических кривых - Малыгина Е.С.
26. Базисы Грёбнера и алгоритм Бухбергера в криптографии 1 - Агиевич С.В.
27. Базисы Грёбнера и алгоритм Бухбергера в криптографии 2 - Агиевич С.В.
28. Группы подстановок в криптографии 1 - Пудовкина М.А.
29. Группы подстановок в криптографии 2 - Пудовкина М.А.
30. Группы подстановок в криптографии 3 - Пудовкина М.А.
31. Группы подстановок в криптографии 4 - Пудовкина М.А.

Студенты успешно окончившие школу:

1. **Аламов Владимир Александрович** – Институт прикладной математики и компьютерных наук, Томский государственный университет, 5 курс специалитета
2. **Атутова Наталья Дмитриевна** – Механико-математический факультет, Новосибирский государственный университет, 2 курс бакалавриата
3. **Базаров Андрей Алдарович** – Факультет информационных технологий, Новосибирский государственный университет, 1 курс бакалавриата
4. **Бахарев Александр Олегович** – Механико-математический факультет, Новосибирский государственный университет, 2 курс бакалавриата
5. **Бонич Татьяна Андреевна** – Механико-математический факультет, Новосибирский государственный университет, 1 курс магистратуры
6. **Быков Денис Александрович** – Механико-математический факультет, Новосибирский государственный университет, 2 курс бакалавриата
7. **Валитов Андрей Александрович** – Факультет информационных технологий, Новосибирский государственный университет, 1 курс бакалавриата
8. **Горяйнова Анастасия Павловна** – Институт математики и компьютерных наук, Тюменский государственный университет, 3 курс бакалавриата
9. **Диденко Андрей Антонович** – Областная специализированная школа лицей для детей одаренных в области математики, физики и информатики, 11 класс, г. Усть-Каменогорск (Казахстан)
10. **Доронин Артемий Евгеньевич** – Механико-математический факультет, Новосибирский государственный университет, 2 курс магистратуры
11. **Дубинская Екатерина Константиновна** – Факультет информационных технологий, Новосибирский государственный университет, 2 курс бакалавриата
12. **Евсюков Анатолий Павлович** – Факультет информационных технологий, Новосибирский государственный университет, 1 курс бакалавриата
13. **Жантуликов Булат Фаритович** – Факультет информационных технологий, Новосибирский государственный университет, 1 курс бакалавриата
14. **Желтова Кристина Анатольевна** – Институт информатики и телекоммуникаций, Сибирский государственный университет науки и технологий им. М.Ф. Решетнева, 3 курс бакалавриата
15. **Завалишина Елена Владимировна** – Факультет информационных технологий, Новосибирский государственный университет
16. **Запольский Максим Михайлович** – Механико-математический факультет, Новосибирский государственный университет, 3 курс бакалавриата
17. **Зюбина Дарья Александровна** – Факультет информационных технологий, Новосибирский государственный университет, 3 курс бакалавриата
18. **Карнаухова Виктория Олеговна** – Факультет информационных технологий, Алтайский государственный технический университет, 1 курс бакалавриата
19. **Касимов Тимур Рустамович** – Факультет прикладной математики и информатики, Новосибирский государственный технический университет, 2 курс бакалавриата
20. **Ким Станислав Евгеньевич** – Факультет прикладной математики и информатики, Новосибирский государственный технический университет, 3 курс бакалавриата
21. **Котельникова Анна Александровна** – Факультет информационных технологий,

Новосибирский государственный университет, 1 курс бакалавриата

22. **Кравец Екатерина Александровна** – Факультет прикладной математики и информатики, Новосибирский государственный технический университет, 2 курс бакалавриата
23. **Крюков Никита Дмитриевич** – Институт прикладной математики и компьютерных наук, Томский государственный университет, 1 курс специалитет
24. **Лаханский Алексей Андреевич** – Факультет информационных технологий, Новосибирский государственный университет, 2 курс бакалавриата
25. **Леонович Дарьяна Александровна** – Факультет прикладной математики и информатики, Новосибирский государственный технический университет, 2 курс бакалавриата
26. **Ляпич Никита Сергеевич** – Муниципальное автономное общеобразовательное учреждение Лицей 6, г. Бердск, выпускник
27. **Матеюк Илья Анатольевич** – Факультет информационных технологий, Новосибирский государственный университет, 2 курс бакалавриата
28. **Натарова Ксения Витальевна** – Факультет радиотехники и кибернетики, Московский физико-технический институт, 1 курс бакалавриата
29. **Никифоров Владислав Сергеевич** – Факультет информационных технологий, Новосибирский государственный университет, 2 курс бакалавриата
30. **Палян Марине Гургеновна** – Факультет информатики и прикладной математики, Ереванский государственный университет, 1 курс магистратуры. Ереван(Армения)
31. **Панферов Матвей Андреевич** – Механико-математический факультет, Новосибирский государственный университет, 1 курс магистратуры
32. **Парфенов Денис Романович** – Факультет информационных технологий, Новосибирский государственный университет, 3 курс бакалавриата
33. **Побединский Сергей Юрьевич** – Факультет прикладной математики и информатики, Новосибирский государственный технический университет, 2 курс бакалавриата
34. **Помыкалов Савелий Витальевич** – Средняя общеобразовательная школа №2, 11 класс, г. Луховицы
35. **Проскурников Никита Андреевич** – Президентский физико-математический лицей 239, 11 класс, г. Санк-Петербург
36. **Разенков Семен Игоревич** – Институт прикладной математики и компьютерных наук, Томский государственный университет, 2 курс специалитета
37. **Раимбеков Азим Русланович** – Институт прикладной математики и компьютерных наук, Томский государственный университет, 3 курс специалитета
38. **Сафенрейтер Дмитрий Алексеевич** – Факультет информационных технологий, Новосибирский государственный университет 2 курс бакалавриата
39. **Семенова Екатерина Вадимовна** – Институт прикладной математики и компьютерных наук, Томский государственный университет, 4 курс специалитета
40. **Сергеев Алексей Вячеславович** – Механико-математический факультет, Новосибирский государственный университет, 2 курс бакалавриата
41. **Сергеев Матвей Игоревич** – Гимназия №1, 10 класс, г. Стерлитамак (респ. Башкортостан)
42. **Скудина Виктория Викторовна** – Механико-математический факультет, Новосибирский государственный университет, 1 курс бакалавриата

43. **Сутормин Иван Александрович** – Механико-математический факультет, Новосибирский государственный университет, 4 курс бакалавриата
44. **Титова Ксения Максимовна** – Механико-математический факультет, Новосибирский государственный университет, 3 курс бакалавриата
45. **Трацевский Игорь Дмитриевич** – Институт математики и компьютерных наук, Тюменский государственный университет, 3 курс специалитета
46. **Хильчук Ирина Сергеевна** – Механико-математический факультет, Новосибирский государственный университет, НГУ, 4 курс бакалавриата
47. **Хлопина София Сергеевна** – Факультет прикладной математики и информатики, Новосибирский государственный технический университет, 1 курс бакалавриата
48. **Ходзицкий Артем Федорович** – Механико-математический факультет, Новосибирский государственный университет, 2 курс бакалавриата
49. **Черников Василий Викторович** – Факультет информационных технологий, Новосибирский государственный университет, 2 курс бакалавриата
50. **Чхайло Иван Дмитриевич** – Институт математики и компьютерных наук, Тюменский государственный университет, 3 курс специалитета
51. **Щербина Даниил Алексеевич** – Институт прикладной математики и компьютерных наук, Томский государственный университет, 3 курс специалитета
52. **Эйсвальд Юлия Игоревна** – Факультет информационных технологий, Новосибирский государственный университет, 1 курс бакалавриата

В 2020 году сотрудниками лаборатории было опубликовано, а также сдано и принято в печать:

- 9 статей в научных журналах
- 7 статей в трудах конференций, из них:
  - 3 статьи в трудах SEIM - Conference on Software Engineering and Information Management, 16 May 2020;
  - 2 статьи в трудах SETA - Sequences and Their Applications, 22-25 September, Russia, Saint-Petersburg. Конференция входит в рейтинг CORE (уровень B);
  - 1 статья в трудах FedCSIS - Conference on computer science and information systems, 6-9 September 2020, Sofia, Bulgaria.
  - 1 статья в трудах CTCrypt – Симпозиум «Современные тенденции в криптографии», 15-17 сентября, Россия, Санкт-Петербург.
- 29 тезисов конференций, из них:
  - 3 тезисов BFA - The 5th International Workshop on Boolean Functions and their Applications, Granada, Spain, September 28 – October 2, 2020;
  - 14 тезисов SIBECRYPT - Сибирская научная школа-семинар с международным участием "Компьютерная безопасность и криптография", 7-11 сентября, г. Томск;
  - 12 тезисов МНСК - Международная научная студенческая конференция, г. Новосибирск, Новосибирский Государственный университет, 10-13 апреля 2020 г.
- Опубликован 100-страничный сборник тезисов летней школы-конференции «Криптография и информационная безопасность» 2020.

  В данный сборник включены 23 работы от 61 автора.

  Сборник доступен по ссылке: http://crypto.nsu.ru/ru/letnyaya-shkola/letnyaya-shkola-2020/

Далее мы приводим тексты самих публикаций.

UDC 519.7 DOI 10.17223/20710410/47/2

A NOTE ON THE PROPERTIES OF ASSOCIATED BOOLEAN FUNCTIONS OF QUADRATIC APN FUNCTIONS[1]

A. A. Gorodilova

Sobolev Institute of Mathematics, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia

E-mail: gorodilova@math.nsc.ru

Let F be a quadratic APN function in n variables. The associated Boolean function γ_F in $2n$ variables ($\gamma_F(a,b)=1$ if $a\neq \mathbf{0}$ and equation $F(x)+F(x+a)=b$ has solutions) has the form $\gamma_F(a,b)=\Phi_F(a)\cdot b+\varphi_F(a)+1$ for appropriate functions $\Phi_F:\mathbb{F}_2^n\to\mathbb{F}_2^n$ and $\varphi_F:\mathbb{F}_2^n\to\mathbb{F}_2$. We summarize the known results and prove new ones regarding properties of Φ_F and φ_F. For instance, we prove that degree of Φ_F is either n or less or equal to $n-2$. Based on computation experiments, we formulate a conjecture that degree of any component function of Φ_F is $n-2$. We show that this conjecture is based on two other conjectures of independent interest.

Keywords: *a quadratic APN function, the associated Boolean function, degree of a function.*

Introduction

Let $\mathbb{F}_2^n$ be the n-dimensional vector space over $\mathbb{F}_2$. Let $\mathbf{0}$ denote the zero vector of $\mathbb{F}_2^n$ and $\mathbf{1}$ denote the vector with all 1s. By «+» we denote the coordinate-wise sum modulo 2 for vectors from $\mathbb{F}_2^n$. Let $x\cdot y=x_1y_1+\ldots+x_ny_n$ denote the *inner product* of vectors $x=(x_1,\ldots,x_n), y=(y_1,\ldots,y_n)\in\mathbb{F}_2^n$; $x\preceq y$ if $x_i\leqslant y_i$ for all $i=1,\ldots,n$; and $\mathrm{wt}(x)=\sum\limits_{i=1}^{n}x_i$ denote the *Hamming weight* of $x\in\mathbb{F}_2^n$. A set $M\subseteq\mathbb{F}_2^n$ forms a *linear subspace* if $x+y\in M$ for any $x,y\in M$; the *dimension* of M, $\dim(M)$, is the maximal number of linearly independent over $\mathbb{F}_2$ vectors from M. We consider *vectorial Boolean functions* $F:\mathbb{F}_2^n\to\mathbb{F}_2^m$, $F=(f_1,\ldots,f_m)$, where $f_i:\mathbb{F}_2^n\to\mathbb{F}_2$, $i=1,\ldots,m$, is a *coordinate* function of F. The *algebraic normal form* (ANF) of F is the following unique representation: $F(x)=\sum\limits_{I\in\mathcal{P}(N)}a_I\left(\prod\limits_{i\in I}x_i\right)$, where $\mathcal{P}(N)$ is the power set of $N=\{1,\ldots,n\}$ and $a_I\in\mathbb{F}_2^m$. The *algebraic degree* of F is degree of its ANF: $\deg(F)=\max\{|I|:\ a_I\neq\mathbf{0},\ I\in\mathcal{P}(N)\}$. Function of algebraic degree at most 1 are called *affine* (they are *linear* in case of $F(\mathbf{0})=\mathbf{0}$). Functions of algebraic degree 2 are called *quadratic*. The *Walsh transform* $W_f:\mathbb{F}_2^n\to\mathbb{Z}$ of a Boolean function $f:\mathbb{F}_2^n\to\mathbb{F}_2$ is defined as $W_f(u)=\sum\limits_{x\in\mathbb{F}_2^n}(-1)^{f(x)+u\cdot x}$. For F the *Walsh spectrum* consists of all *Walsh coefficients* $W_{F_v}(u)$, $u\in\mathbb{F}_2^n$, $v\in\mathbb{F}_2^m$, $v\neq\mathbf{0}$, where $F_v=v\cdot F$ is a *component* Boolean function of F.

A function F from $\mathbb{F}_2^n$ to itself is called *almost perfect nonlinear* (APN) (according to K. Nyberg [1]) if for any $a,b\in\mathbb{F}_2^n$, $a\neq\mathbf{0}$, equation $F(x)+F(x+a)=b$ has at most two solutions. APN functions are of special interest for using as S-boxes in block ciphers

[1]The work was funded by RFBR (projects no. 18-31-00479, 18-07-01394); by the program of fundamental scientific researches of the SB RAS no. I.5.1, project no. 0314-2019-0017; Regional Mathematical Center NSU and Laboratory of cryptography JetBrains Research.

due to their optimal differential characteristics. Despite the fact that APN functions are intensively studied (see, for example, the book of L. Budaghyan [2], surveys of A. Pott [3], M. M. Glukhov [4], and M. E. Tuzhilin [5]), there are a lot of open problems on finding new constructions, classifications, etc.

In [6], C. Carlet, P. Charpin, and V. Zinoviev introduced the *associated Boolean function* $\gamma_F : \mathbb{F}_2^{2n} \to \mathbb{F}_2$ for a given vectorial Boolean function F from $\mathbb{F}_2^n$ to itself; $\gamma_F(a, b) = 1$ if and only if $a \neq \mathbf{0}$ and equation $F(x) + F(x + a) = b$ has solutions.

Two functions are called *differentially equivalent* [7] (or γ-equivalent according to K. Boura et al. [8]), if their associated Boolean functions coincide. The problem of describing the differential equivalence class of an APN function remains open even for quadratic case. That is why we are interested in obtaining some properties of γ_F. We will focus on quadratic APN functions.

Let F be a quadratic APN function. Then the set $B_a(F) = \{F(x) + F(x + a) : x \in \mathbb{F}_2^n\}$ is a linear subspace of dimension $n - 1$ or its complement for a nonzero $a \in \mathbb{F}_2^n$. Using this fact, γ_F can be uniquely represented in the form

$$\gamma_F(a, b) = \Phi_F(a) \cdot b + \varphi_F(a) + 1,$$

where $\Phi_F : \mathbb{F}_2^n \to \mathbb{F}_2^n$, $\varphi_F : \mathbb{F}_2^n \to \mathbb{F}_2$ are defined from $B_a(F) = \{y \in \mathbb{F}_2^n : \ \Phi_F(a) \cdot y = \varphi_F(a)\}$ for all $a \neq \mathbf{0}$; and $\Phi_F(\mathbf{0}) = \mathbf{0}$, $\varphi_F(\mathbf{0}) = 1$. Note that $B_a(F)$ is a linear subspace if and only if $\varphi_F(a) = 0$. It is easy to see that $(F(x) + F(x + a) + F(a) + F(\mathbf{0})) \cdot \Phi_F(a) = \mathbf{0}$ for all $x \in \mathbb{F}_2^n$ by definition.

In this paper we study the properties of functions Φ_F and φ_F.

1. Properties of φ_F and Φ_F

In this section we summarize known results and present new ones about properties of Φ_F and φ_F. As it usually happens the cases of even and odd number of variables are different.

1.1. The image set of Φ_F

According to [9], let us denote $A_v^F = \{a \in \mathbb{F}_2^n : \Phi_F(a) = v\}$.

Theorem 1 [6, 9]. Let F be a quadratic APN function in n variables.

1) If n is odd, then Φ_F is a permutation.
2) If n is even, then the preimage Φ_F of any nonzero vector is a linear subspace of even dimension together with the zero vector.

Note that state 1 in Theorem 1 means also that γ_F is a bent function of Maiorana—McFarland type (readers may find details regarding bent functions in [10]).

Corollary 1. Let F be a quadratic APN function. Then Φ_F takes an odd number of distinct nonzero values.

Proof. By definition of Φ_F, we have $\Phi_F(\mathbf{0}) = \mathbf{0}$.

If n is odd, then Φ_F is a permutation [6]. Hence, the proposition holds.

Let n be even. It is known [9] that the preimage set $A_v^F = \{x \in \mathbb{F}_2^n : \Phi_F(x) = v\}$ for any nonzero $v \in \mathbb{F}_2^n$ represents a linear subspace of even dimension together with the zero vector. Let $\Phi_F \in \{\mathbf{0}, v_1, \ldots, v_m\}$, where v_i, $i = 1, \ldots, m$, are pairwise distinct nonzero vectors. We need to prove that m is odd. We have

$$2^n - 1 = |A_{v_1}^F| + \ldots + |A_{v_m}^F| = 2^{\lambda_1} - 1 + \ldots + 2^{\lambda_m} - 1 = 2^{\lambda_1} + \ldots + 2^{\lambda_m} - m,$$

where λ_i, $i = 1, \ldots, m$, are nonzero even numbers. Since $2^n - 1$ is odd, then m is also odd. ■

1.2. The degree of φ_F

Proposition 1. Let F be a quadratic APN function in n variables, n is even. Then $\deg(\varphi_F) = n$, or, equivalently, $\mathrm{wt}(\varphi_F)$ is odd.

Proof. It is known [9] that $A_v^F \cup \{\mathbf{0}\}$ is a linear subspace of even dimension if n is even for any nonzero $v \in \mathbb{F}_2^n$. Also [9], there exists $c_v \in \mathbb{F}_2^n$ such that $\varphi_F|_{A_v^F} = c_v \cdot x|_{A_v^F}$. Hence, $\mathrm{wt}(\varphi_F|_{A_v^F})$ is an even number equal to 0 or $2^{\dim(A_v^F \cup \{\mathbf{0}\})-1}$ for any nonzero v and $\varphi_F(\mathbf{0}) = 1$ by definition. Thus, $\mathrm{wt}(\varphi_F)$ is odd. It is widely known that $\mathrm{wt}(f)$ is odd if and only if $\deg(f) = n$ for any Boolean function in n variables. ■

The case of odd n remains open. Based on our computational experiments for all known quadratic APN functions of not more than 11 variables, we can formulate the following

Conjecture 1. Let F be a quadratic APN function in n variables, n is odd. Then $\deg(\varphi_F) < n$, or, equivalently, $\mathrm{wt}(\varphi_F)$ is even.

1.3. The degree of Φ_F

Theorem 2 [7]. Let F be a quadratic APN function in n variables, $n \geqslant 3$, n is odd. Then $\deg(\Phi_F) \leqslant n - 2$.

The following theorem contains a similar bound for even n.

Theorem 3. Let F be a quadratic APN function in n variables, $n \geqslant 4$, n is even. Then each coordinate function of Φ_F is represented as $(\Phi_F)_i(x) = f_i(x) + \lambda_i\big(x_2 \dots x_n + + x_1x_3 \dots x_n + \dots + x_1x_2 \dots x_{n-1} + x_1 \dots x_n\big)$, where $\deg(f_i) \leqslant n - 2$ and $\lambda_i \in \mathbb{F}_2$.

Proof. Let $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a linear function. Then it is easy to see that

$$\gamma_{F+L}(a, b) = \gamma_F(a, b+L(a)) = (b+L(a))\cdot\Phi_F(a)+\varphi_F(a)+1 = b\cdot\Phi_F(a)+\varphi_F(a)+L(a)\cdot\Phi_F(a)+1.$$

Hence, $\Phi_{F+L} = \Phi_F$ and $\varphi_{F+L} = \varphi_F + L \cdot \Phi_F$. By Proposition 1, $\deg(\varphi_F) = \deg(\varphi_{F+L}) = n$, since $F+L$ is also a quadratic APN function. Thus, $\deg(L\cdot\Phi_F) < n$ for any linear function L.

Suppose that $\deg(\Phi_F) = n$. This means that there exists a coordinate function $(\Phi_F)_i$ of degree n. Let us represent

$$(\Phi_F)_i(x) = f_i(x) + a_1x_2 \dots x_n + a_2x_1x_3 \dots x_n + \dots + a_nx_1x_2 \dots x_{n-1} + x_1 \dots x_n,$$

where $\deg(f_i) \leqslant n - 2$ and $a_1, \dots, a_n \in \mathbb{F}_2$.

— If $a_j = 0$, then $\deg(L \cdot \Phi_F) = n$ for $L = (0, \dots, 0, x_j, 0, \dots, 0)$, where x_j is the i-th coordinate function of L. Hence, we get a contradiction.
— If $a_j = 1$ for all j, then it is easy to see that we will always have $\deg(L \cdot \Phi_F) < n$ for any linear function L.

Suppose that $\deg(\Phi_F) = n - 1$. Similarly,

$$(\Phi_F)_i(x) = f_i(x) + a_1x_2 \dots x_n + a_2x_1x_3 \dots x_n + \dots + a_nx_1x_2 \dots x_{n-1},$$

where at least one coefficient is equal to 1, say a_j. Then $\deg(L \cdot \Phi_F) = n$ for $L = (0, \dots, 0, x_j, 0, \dots, 0)$, where x_j is the i-th coordinate function of L. Hence, we get a contradiction.

Thus, $(\Phi_F)_i$ is of degree not more than $n - 2$ or all monomials of degree $n - 1$ and n are included in the ANF of $(\Phi_F)_i$. ■

Remark 1. For all known quadratic APN functions in not more than 11 variables, we computationally verified that

— for even n, the case $\deg((\Phi_F)_i) = n$ is not realized;

— any component function of Φ_F has degree exactly $n-2$.

Based on our computational experiments we can formulate the following

Conjecture 2. Let F be a quadratic APN function in n variables, $n \geqslant 3$. Then $\deg(v \cdot \Phi_F) = n-2$ for any nonzero $v \in \mathbb{F}_2^n$.

2. Does the equality $\deg(\Phi_F) = n-2$ hold?

In this section we study the following question: "Is conjecture 2 true or not?".

For example, consider an APN Gold function $F(x) = x^{2^k+1}$, $\gcd(n,k) = 1$ (the function is given as a function over the finite field of order 2^n). Its associated Boolean function is known [6]: $\gamma_F(a,b) = \mathrm{tr}((a^{2^k+1})^{-1}b) + \mathrm{tr}(1) + 1$ (here tr is the absolute trace function in the finite field of order 2^n). So, we have $\Phi_F(a) = (a^{2^k+1})^{-1}$, $\Phi_F(0) = 0$, and as it is easy to see $\deg(\Phi_F) = n-2$ (since it is well-known that the degree of a function $F(x) = x^d$ is equal to the 2-weight of the integer d modulo 2^n).

We wonder whether conjecture 2 is true or not for arbitrary n. Let us focus on the case of odd n since in this case we have the bound of Theorem 2. For even case, the consideration could be rather similar but with assumption that $\deg(\Phi_F)$ is not equal to n, that is only a conjecture up to now.

S t e p 1. Let F be a quadratic APN function of n variables, n is odd, $n \geqslant 5$; v be a nonzero vector from $\mathbb{F}_2^n$. We need to prove that $\deg(v \cdot \Phi_F) = n-2$ for any nonzero $v \in \mathbb{F}_2^n$.

We use the following widely known equality for counting the ANF coefficients of a Boolean function f in n variables:

$$g_f(a) = \Big(2^{\mathrm{wt}(a)-1} - 2^{\mathrm{wt}(a)-n-1} \sum_{b \preceq (a+\mathbf{1})} W_f(b)\Big) \bmod 2. \tag{1}$$

We need to show that there exists a vector a^v with $\mathrm{wt}(a^v) = n-2$ such that $g_{v\cdot\Phi_F}(a^v) = 1$. Equivalently, that there exist coordinates i, j, $1 \leqslant i \neq j \leqslant n$, such that

$$\sum_{b \preceq (a^v+\mathbf{1})} W_{v\cdot\Phi_F}(b) = W_{v\cdot\Phi_F}(\mathbf{0}) + W_{v\cdot\Phi_F}(e^i) + W_{v\cdot\Phi_F}(e^j) + W_{v\cdot\Phi_F}(e^i+e^j)$$

is not divided by 16 according to (1). Here e^i is the vector with 1 in the i-th coordinate and 0s in other coordinates. Let us introduce the following sets:

$$M^i = \{x \in \mathbb{F}_2^n : v \cdot \Phi_F(x) = 0,\ x \cdot e^i = 0\},$$
$$M^j = \{x \in \mathbb{F}_2^n : v \cdot \Phi_F(x) = 0,\ x \cdot e^j = 0\},$$
$$M^{ij} = \{x \in \mathbb{F}_2^n : v \cdot \Phi_F(x) = 0,\ x \cdot (e^i+e^j) = 0\}.$$

Then, we have

$$\sum_{b \preceq (a^v+\mathbf{1})} W_{v\cdot\Phi_F}(b) = 4|M^i| - 2^n + 4|M^j| - 2^n + 4|M^{ij}| - 2^n =$$
$$= 4(|M^i| + |M^j| + |M^{ij}|) - 3\cdot 2^n = 4(2^{n-1} + 2|M_0^{ij}|) - 3\cdot 2^n = 8|M_0^{ij}| - 2^{n-1},$$

where $M_0^{ij} = \{x \in \mathbb{F}_2^n : v \cdot \Phi_F(x) = 0,\ x \cdot e^i = 0,\ x \cdot e^j = 0\}$.

S t e p 2. Thus, we need to prove that there exist coordinates i, j, $1 \leqslant i \neq j \leqslant n$, such that $|M_0^{ij}|$ is odd (since we consider $n \geqslant 5$). From [7, prop. 7], we know that $M = \{x \in \mathbb{F}_2^n : v \cdot \Phi_F(x) = 0\} = \bigcup\limits_{\ell \in I} A_\ell$, where A_ℓ is a linear subspace of dimension 2,

and $A_\ell \cap A_k = \{\mathbf{0}\}$, $\ell, k \in I$, $\ell \neq k$. Since Φ_F is a permutation, then $|M| = 2^{n-1}$ and $|I| = (2^{n-1}-1)/3$.

Let us consider an arbitrary $A_\ell = \{\mathbf{0}, x^\ell, y^\ell, x^\ell + y^\ell\}$. Then for any distinct coordinates i, j of $x^\ell, y^\ell, x^\ell + y^\ell$ we have the following situations (without permutations of rows):

	ij		ij		ij		ij		ij
x^ℓ	00		00		00		00		01
y^ℓ	00	or	01	or	10	or	11	or	10
$x^\ell + y^\ell$	00		01		10		11		11

Hence, the number of $x^\ell, y^\ell, x^\ell + y^\ell$ together with $\mathbf{0}$ that belong to the set M_0^{ij} is equal to $1+3\cdot N_3^{ij}+1\cdot N_1^{ij}+0\cdot N_0^{ij}$, where $N_3^{ij}+N_1^{ij}+N_0^{ij} = |I| = (2^{n-1}-1)/3$, and N_k^{ij}, $k = 0, 1, 3$, is the number of A_ℓ, $\ell \in I$, having exactly k vectors with both coordinates i and j equal to 0. Thus, $|M_0^{ij}|$ is odd if and only if N_0^{ij} is odd.

S t e p 3. Now, we need to prove that there exist coordinates i, j, $1 \leqslant i \neq j \leqslant n$, such that N_0^{ij} is odd. We found the following interesting property (computationally verified for $n = 5$) that we formulate as a conjecture.

Conjecture 3. Let $M = \bigcup\limits_{\ell \in I} A_\ell$, where A_ℓ is a linear subspace of dimension 2, and $A_\ell \cap A_k = \{\mathbf{0}\}$, $\ell, k \in I$, $\ell \neq k$, $|I| = (2^{n-1}-1)/3$. Then the set M is a hyperplane $\{x \in \mathbb{F}_2^n : x_m = 0\}$ for some coordinate m if and only if the number of subspaces A_ℓ without elements having both coordinates i and j equal to 0 is even for any distinct coordinates i, j.

S t e p 4. If Conjecture 3 is true, then we need to prove that $M = \{x \in \mathbb{F}_2^n : v \cdot \Phi_F(x) = 0\}$ cannot be a hyperplane $\{x \in \mathbb{F}_2^n : x_m = 0\}$ for some coordinate m.

Conjecture 4. Let F be a quadratic APN function in n variables, $n \geqslant 5$. Then $\{x \in \mathbb{F}_2^n : v \cdot \Phi_F(x) = 0\}$ is not a linear subspace.

We computationally verified this property for all known quadratic APN functions for $n = 5, \ldots, 11$ and formulate the conjecture.

Thus, by proving Conjectures 3 and 4, we can prove the starting Conjecture 2. Unfortunately, each of them remains open up to now.

Conclusion

The following question is open: what properties must a Boolean function satisfy in order to be the associated function for some vectorial function? Even a partial answer to the question provides a potential method to find new APN functions if we can choose "admissible" Boolean functions as γ_F. For example, using the algorithm from [8] for reconstructing APN functions from its associated functions. Another reason why we study the properties of associated functions is that they may lead to new results in the differential equivalence classification of APN functions.

REFERENCES

1. *Nyberg K.* Differentially uniform mappings for cryptography. Advances in Cryptography, EUROCRYPT'93, LNCS, 1994, vol. 765, pp. 55–64.
2. *Budaghyan L.* Construction and Analysis of Cryptographic Functions. Springer International Publishing, 2014. 168 p.
3. *Pott A.* Almost perfect and planar functions. Designs, Codes and Cryptography, 2016, vol. 78, pp. 141–195.
4. *Glukhov M. M.* O priblizhenii diskretnykh funktsiy lineynymi funktsiyami [On the approximation of discrete functions by linear functions]. Matematicheskie Voprosy Kriptografii, 2016, vol. 7, no. 4, pp. 29–50. (in Russian)

5. *Tuzhilin M. E.* Pochti sovershennye nelineynye funktsii [APN-functions]. Prikladnaya Diskretnaya Matematika, 2009, no. 3 (5), pp. 14–20. (in Russian)
6. *Carlet C., Charpin P., and Zinoviev V.* Codes, bent functions and permutations suitable for DES-like cryptosystems. Designs, Codes and Cryptography, 1998, vol. 15, iss. 2, pp. 125–156.
7. *Gorodilova A.* On the differential equivalence of APN functions. Cryptography and Communications, 2019, vol. 11, iss. 4, pp. 793–813.
8. *Boura C., Canteaut A., Jean J., and Suder V.* Two notions of differential equivalence on Sboxes. Designs, Codes and Cryptography, 2019, vol. 87, iss. 2–3, pp. 185–202.
9. *Gorodilova A.* Lineynyy spektr kvadratichnykh APN-funktsiy [The linear spectrum of quadratic APN functions]. Prikladnaya Diskretnaya Matematika, 2016, no 4(34), pp. 5–16. (in Russian)
10. *Tokareva N.* Bent Functions, Results and Applications to Cryptography. Acad. Press. Elsevier, 2015. 230 p.

On the Sixth International Olympiad in Cryptography NSUCRYPTO*

A. Gorodilova[1], N. Tokareva[1,2], S. Agievich[3], C. Carlet[4], E. Gorkunov[1,5], V. Idrisova[1], N. Kolomeec[1], A. Kutsenko[1,5], R. Lebedev[5], S. Nikova[6], A. Oblaukhov[1], I. Pankratova[7], M. Pudovkina[8], V. Rijmen[6], A. Udovenko[9]

[1]Sobolev Institute of Mathematics, Novosibirsk, Russia
[2]Laboratory of Cryptography JetBrains Research
[3]Belarusian State University, Minsk, Belarus
[4]University of Paris 8, Paris, France
[5]Novosibirsk State University, Novosibirsk, Russia
[6]ESAT-COSIC, KU Leuven, Leuven, Belgium
[7]Tomsk State University, Tomsk, Russia
[8]Bauman Moscow State Technical University, Moscow, Russia
[9]SnT, University of Luxembourg, Esch-sur-Alzette, Luxembourg

E-mail: nsucrypto@nsu.ru

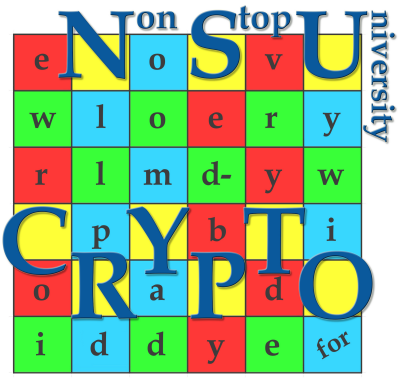
Abstract. NSUCRYPTO is the unique cryptographic Olympiad containing scientific mathematical problems for professionals, school and university students from any country. Its aim is to involve young researchers in solving curious and tough scientific problems of modern cryptography. From the very beginning, the concept of the Olympiad was not to focus on solving olympic tasks but on including unsolved research problems at the intersection of mathematics and cryptography. The Olympiad history starts in 2014. In 2019, it was held for the sixth time. In this paper, problems and their solutions of the Sixth International Olympiad in cryptography NSUCRYPTO'2019 are presented. We consider problems related to attacks on ciphers and hash functions, protocols, Boolean functions, Dickson polynomials, prime numbers, rotor machines, etc. We discuss several open problems on mathematical countermeasures to side-channel attacks, APN involutions, S-boxes, etc. The problem of finding a collision for the hash function `Curl27` was partially solved during the Olympiad.

Keywords. cryptography, ciphers, hash functions, Hamming code, slide attack, threshold implementation, Dickson polynomial, APN function, Olympiad, NSUCRYPTO.

*The work of the first two authors and the sixth author was supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research. The work of the seventh, eighth and eleventh authors was supported by Russian Foundation for Basic Research (projects no. 20-31-70043, 18-07-01394, 19-31-90093).

Introduction

NSUCRYPTO (Non-Stop University Crypto) is the International Olympiad in cryptography that was held for the sixth time in 2019.



Interest in the Olympiad around the world is significant. This year, there were hundreds of participants from 26 countries; 42 participants in the first round and 21 teams in the second round from 16 countries were awarded with prizes and honorable diplomas. The Olympiad program committee includes specialists from Belgium, France, the Netherlands, the USA, Norway, India, Luxembourg, Belarus', Kazakhstan, and Russia.

Let us shortly formulate the format of the Olympiad. One of the Olympiad main ideas is that everyone can participate! Each participant chooses his/her category when registering on the Olympiad website [15]. There are three categories: "school students" (for junior researchers: pupils and high school students), "university students" (for participants who are currently studying at universities) and "professionals" (for participants who have already completed education or just want to be in the restriction-free category). Awarding of the winners is held in each category separately.

The Olympiad consists of two independent Internet rounds: the first one is individual (duration 4 hours 30 minutes) while the second round is a team one (duration 1 week). The first round is divided into two sections: A — for "school students", B — for "university students" and "professionals". The second round is common to all participants. Participants read the Olympiad problems and submit their solutions using the Olympiad website. The language of the Olympiad is English.

The Olympiad participants are always interested in solving different problems of various complexities at the intersection of mathematics and cryptography. They show their knowledge, creativity and professionalism. That is why the Olympiad not only includes interesting tasks with known solutions but also offers unsolved problems in this area. This year, one of such open problems, "`Curl27`" (see section 2.14), was partially solved during the second round! All the open problems stated during the Olympiad history can be found here [16]. On the website we also mark the current status of each problem. For example, in addition to "`Curl27`", the problem "Sylvester matrices" was solved by three teams in 2018, the problem "Algebraic immunity" was completely solved during the Olympiad in 2016. And what is important for us, some participants were trying to find solutions after the Olympiad was over. For example, a partial solution for the problem "A secret sharing" (2014) was proposed in [9]. We invite everybody who has ideas on how to solve the problems to send your solutions to us!

The paper is organized as follows. We start with problem structure of the Olympiad in section 1. Then we present formulations of all the problems stated during the Olympiad and give their detailed solutions in section 2. Finally, we publish the lists of NSUCRYPTO'2019 winners in section 3.

Mathematical problems and their solutions of the previous International Olympiads in cryptography NSUCRYPTO from 2014 to 2018 can be found in [2], [1], [14], [10], and [11] respectively.

1 Problem structure of the Olympiad

There were 16 problems stated during the Olympiad, some of them were included in both rounds (Tables 1, 2). Section A of the first round consisted of six problems, whereas the section B contained

seven problems. Three problems were common for both sections. The second round was composed of eleven problems. Five problems of the second round included unsolved questions (awarded special prizes from the Program Committee).

Table 1: **Problems of the first round**

N	Problem title	Maximum scores
1	A 1024-bit key	4
2	The magnetic storm	4
3	Autumn leaves	4
4	A rotor machine	4
5	Broken `Calculator`	4
6	A promise	6

Section A

N	Problem title	Maximum scores
1	Autumn leaves	4
2	The magnetic storm	4
3	A rotor machine	4
4	16QAM	8
5	A promise and money	6
6	`Calculator`	6
7	APN + Involutions	7

Section B

Table 2: **Problems of the second round**

N	Problem title	Maximum scores
1	A 1024-bit key	4
2	Sharing	6 + additional scores for open questions
3	Factoring in 2019	8
4	`TwinPeaks-3`	8
5	`Curl27`	10 + additional scores for open questions
6	8-bit S-box	Unlimited (open problem)
7	A rotor machine	4
8	16QAM	8
9	`Calculator`	6
10	APN + Involutions (extended)	12 + additional scores for open questions
11	Conjecture	Unlimited (open problem)

2 Problems and their solutions

In this section, we formulate all the problems of NSUCRYPTO'2019 and present their detailed solutions paying attention to solutions proposed by the participants.

2.1 Problem "A 1024-bit key"

2.1.1 Formulation

Alice has a 1024-bit key for a symmetric cipher (the key consists of 0s and 1s). Alice is afraid of malefactors, so she changes her key everyday in the following way:

1. Alice chooses a subsequence of key bits such that the first bit and the last bit are equal to 0. She also can choose a subsequence of length 1 that contains only 0.
2. Alice inverts all the bits in this subsequence (0 turns into 1 and vice versa); bits outside of this subsequence remain as they are.

Prove that the process will stop. Find the key that will be obtained by Alice in the end of the process.

Example of an operation. $11001\underbrace{01101110}011...$ turns to $11001\underbrace{10010001}011...$

2.1.2 Solution

Let us encode the binary vector of the key as the corresponding decimal number. It is obvious that this number will increase on the next day, since all the bits on the left from the sequence are not changing, but the first bit of the sequence turns from 0 to 1. Let us note that this number can not increase infinitely since the size of the key is restricted by 1024 bits, so, in the very end the key will be maximal possible and, thus, will consist of all 1s.

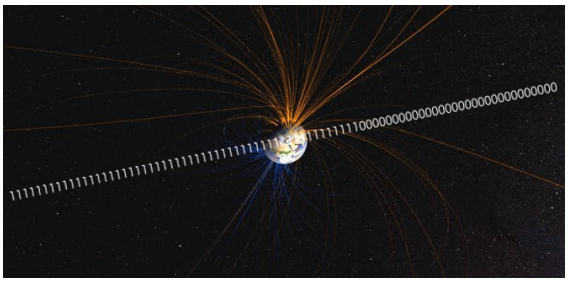
Almost all the participants successfully solved the problem.

2.2 Problem "The magnetic storm"

2.2.1 Formulation

A hardware random number generator is a device that generates random sequences consisting of 0s and 1s. Unfortunately, a disturbance caused by a magnetic storm affected this random number generator. As a result, the device had generated a sequence of 0s of length k (where k is a positive integer), and then started to generate an infinite sequence of 1s.

Prove that at some point the generator will produce a number $1\ldots10\ldots0$ that is divisible by 2019.



2.2.2 Solution

Let us prove that a number of form $1\ldots11\ldots1$ is divisible by 2019. Consider all numbers that consists only of 1s, since there are infinite quantity of these numbers, there can be found a pair of numbers A and B such that they have the same remainder when divided by 2019. Therefore, $C = A - B = 1\ldots10\ldots0$ consisting of m 1s for some natural m is divisible by 2019, and, since 2019 is not divisible by 2 and 5, $C^* = C \times 10\ldots0 = 1\ldots10\ldots0$ is divisible by 2019 for any number of 0s.

There were a lot of correct solutions from the participants.

2.3 Problem “Autumn leaves”

2.3.1 Formulation

Read a hidden message!..

2.3.2 Solution

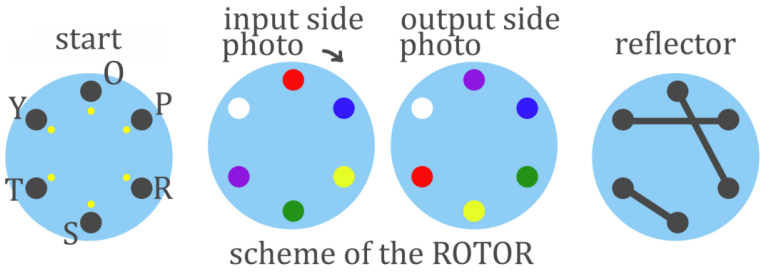
We see different leaves and spaces between them. It looks like a simple substitution cipher was used there and distinct leaves corresponded to distinct English letters. By English grammar, we can suppose that the second and the third words are “`is a`”. Then the first word starts with “`a`” and by its structure can be “`autumn`” (which is very likely as the autumn landscape is depicted). Also, the leaf  is the most common letter in the text and we can guess that it is “`e`”. Then we see “`*ea*`” in the third line that seems to be “`leaf`”. As a result the last word becomes “`fl**e*`” that is “`flower`”. Finally, we get “`Autumn is a second spring when every leaf is a flower`” that is a famous quote by Albert Camus. Almost all the participants read the message.

2.4 Problem “A rotor machine”

2.4.1 Formulation

In one country rotor machines were very useful for encryption of information.

Eve knows that for some secret communication a simple rotor machine was used. It works with letters `O, P, R, S, T, Y` only and has an input circle with lamps (start), one rotor and a reflector. See the scheme below.



The input circle and the reflector are fixed in their positions while the rotor can be in one of 6 possible positions. After pressing a button on a keyboard, an electrical signal corresponding to the letter goes through the machine, comes back to the input circle, and the appropriate lamp shows the result of encryption. After each letter is encrypted, the rotor turns right (i. e. clockwise) on 60 degrees. Points of different colors on the rotor sides indicate different noncrossing signal lines within the rotor.

For instance, if the rotor is fixed as shown on the picture above, then if you press the button `O`, it will be encrypted as `T` (the signal enters the rotor via red point, is reflected and then comes back via purple line). If you press `O` again, it will be encrypted as `R`. If you press `T` then, you will get `S` and so on.

Eve intercepted a secret message: `TRRYSSPRYRYROYTOPTOPTSPSPRS`. Help her to decrypt it keeping in mind that Eve does not know the initial position of the rotor.

2.4.2 Solution

To solve the problem and decrypt the message, one needs to correctly understand the scheme of work. A key for the cipher is the initial position of the rotor. We denote it by a color of the circle on the input side of the rotor that corresponds to the letter `O`. Table 3 represents the encryption tables depending on the key.

Trying all six possible keys, we find the only one meaningful message `POST TO TOP OOPS SORRY STOP ROTOR` that corresponds to the “yellow” key.

Almost all the participants solved the problem. The most interesting solutions were obtained by creating real models for this rotor machine, for example by a school student Varvara Lebedinskaya

Table 3: **Encryption tables**

	O	P	R	S	T	Y
red	T	Y	S	R	O	P
white	R	S	O	P	Y	T
purple	Y	R	P	T	S	O

	O	P	R	S	T	Y
green	S	R	P	O	Y	T
yellow	S	T	Y	O	P	R
blue	R	T	O	Y	P	S

(The Specialized Educational Scientific Center of Novosibirsk State University), by the team of Kristina Geut, Sergey Titov, and Dmitry Ananichev (Ural State University of Railway Transport).

2.5 Problem "Broken `Calculator`"

2.5.1 Formulation

Alice and Bob are practicing in developing toy cryptographic applications for smartphones. This year they have invented `Calculator` that allows one to perform the following operations modulo 2019 (that is to get the result as the reminder of division by 2019):

- to insert at most 4-digit positive integers (digits from 0 to 9);
- to perform addition, subtraction and multiplication of two numbers;
- to store temporary results and read them from the memory.

Suppose that Alice wants to send Bob a ciphertext y (given by a 4-digit integer). She sends y from her smartphone to Bob's `Calculator` memory. To decrypt y, Bob needs to get the plaintext x (using his `Calculator`) by the rule: x is equal to the remainder of dividing $f(y) = y^5 + 1909y^3 + 401y$ by 2019.

At the most inopportune moment, Bob dropped his smartphone and broke its screen Now, the button [+] as well as all digits except [1] and [5] are not working.

Help Bob to invent an efficient algorithm how to decrypt any ciphertext y using `Calculator` in his situation. More precisely, suggest a short list of commands, where each command has one of the following types ($1 \leqslant j, k < i$):

$$S_i = y, \qquad S_i = a, \qquad S_i = S_j - S_k, \qquad S_i = S_j * S_k,$$

where a is an at most 4-digit integer consisting of digits 1 and 5 only; for example, $a = 1$, $a = 15$, $a = 551$, $a = 5115$, etc.

The first command has to be $S_1 = y$. In the last command, the resulting plaintext x has to be calculated. We remind that all calculations are modulo 2019. In particular, the integer 2500 becomes 481 and -1000 becomes 1019 immediately after entering or calculations. The shorter the list of commands you suggest, the more scores you get for this problem.

Example. The following list of commands calculates $x = y^2 - 55$:

Command	Result
$S_1 = y$	y
$S_2 = S_1 * S_1$	y^2
$S_3 = 11$	11
$S_4 = 5$	5
$S_5 = S_3 * S_4$	55
$S_6 = S_2 - S_5$	$y^2 - 55$

2.5.2 Solution

Let us present the original solution by the programm committee that has 14 steps.

Let $a \equiv_m b$ mean that integers a and b are congruent modulo m. The following relations hold:

$$\begin{aligned} f(y) &\equiv_{2019} y^5 + 1909y^3 + 401y \\ &\equiv_{2019} y(y^4 - 110y^2 + 401) \\ &\equiv_{2019} y(y^4 - 2 * 55y^2 + 55^2 - 55^2 + 401) \\ &\equiv_{2019} y((y^2 - 55)^2 - 55^2 + 5 * 22^2) \\ &\equiv_{2019} y((y^2 - 55)^2 - 11^2 * (5^2 - 5 * 2^2)) \\ &\equiv_{2019} y((y^2 - 55)^2 - 11^2 * 5) \\ &\equiv_{2019} y((y^2 - 55)^2 - 11 * 55). \end{aligned}$$

Thus, the reminder of division of $f(y)$ by 2019 can be calculated for any y by the list of commands given in Table 4. A similar solution was found by Borislav Kirilov (Bulgaria, The First Private Mathematical Gymnasium).

Table 4: **List of commands for the programm committee solution**

Command	Result	Command	Result	Command	Result
$S_1 = y$	y	$S_4 = S_2 - S_3$	$y^2 - 55$	$S_7 = S_3 * S_6$	$11 * 55$
$S_2 = S_1 * S_1$	y^2	$S_5 = S_4 * S_4$	$(y^2 - 55)^2$	$S_8 = S_5 - S_7$	$(y^2 - 55)^2 - 11 * 55$
$S_3 = 55$	55	$S_6 = 11$	11	$S_9 = S_1 * S_8$	$y((y^2 - 55)^2 - 11 * 55)$

Note. The polynomial $f(y) = y^5 + 1909y^3 + 401y$ is the Dickson polynomial $D_5(y, a) = y^5 - 5y^3a + 5ya^2$ for $a = 22$ with coefficients taken modulo 2019.

2.6 Problem "Calculator"

2.6.1 Formulation

Alice and Bob are practicing in developing toy cryptographic applications for smartphones. This year they have invented `Calculator` that allows one to perform the following operations modulo 2019:

- to insert at most 4-digit positive integers (digits from 0 to 9);
- to perform addition, subtraction and multiplication of two numbers;
- to store temporary results and read them from the memory.

Suppose that Alice wants to send Bob a ciphertext y (given by a 4-digit integer). She sends y from her smartphone to Bob's `Calculator` memory. To decrypt y, Bob needs to get the plaintext x (using his `Calculator`) by the rule $x = f(y) \bmod 2019$, where f is a secret polynomial known to Alice and Bob only.

At the most inopportune moment, Bob dropped his smartphone and broke its screen Now, the button [+] as well as all digits except [2] are not working.

Help Bob to invent an efficient algorithm how to decrypt any ciphertext y using `Calculator` in his situation if the current secret polynomial is $f(y) = y^5 + 1909y^3 + 401y$. More precisely, suggest a short list of commands, where each command has one of the following types ($1 \leqslant j, k < i$):

$$S_i = y, \qquad S_i = 2, \qquad S_i = 22, \qquad S_i = 222, \qquad S_i = 2222, \qquad S_i = S_j - S_k, \qquad S_i = S_j * S_k.$$

The first command has to be $S_1 = y$. In the last command, the resulted plaintext x has to be calculated. We remind that all calculations are modulo 2019. In particular, the integer 2222 becomes 203 immediately after entering. The shorter the list of commands you suggest, the more scores you get for this problem.

Example. The following list of commands calculates $x = y^2 - 4$:

Command	Result
$S_1 = y$	y
$S_2 = S_1 * S_1$	y^2
$S_3 = 2$	2
$S_4 = S_3 * S_3$	4
$S_5 = S_2 - S_4$	$y^2 - 4$

2.6.2 Solution

The polynomial $f(y) = y^5 + 1909y^3 + 401y$ is the Dickson polynomial $D_5(y, a) = y^5 - 5y^3a + 5ya^2$ for $a = 22$ with coefficients taken modulo 2019. The following relations hold:

$$\begin{aligned} D_5(y,a) &= yD_4(y,a) - aD_3(y,a) \\ &= yD_2(D_2(y,a), a^2) - aD_3(y,a) \\ &= y((y^2-2a)^2 - 2a^2) - ay(y^2 - 2a - a). \end{aligned}$$

For $a = 22$, the value $f(y)$ can be calculated for any y by the list of commands given in Table 5.

Table 5: **List of commands for the programm committee solution**

Command	Result	Command	Result
$S_1 = y$	y	$S_8 = S_7 * S_7$	$(y^2 - 2a)^2$
$S_2 = 2$	2	$S_9 = S_8 - S_5$	$(y^2 - 2a)^2 - 2a^2$
$S_3 = 22$	a	$S_{10} = S_1 * S_9$	$y((y^2 - 2a)^2 - 2a^2)$
$S_4 = S_2 * S_3$	$2a$	$S_{11} = S_7 - S_2$	$y^2 - 2a - a$
$S_5 = S_3 * S_4$	$2a^2$	$S_{12} = S_1 * S_{11}$	$y(y^2 - 2a - a)$
$S_6 = S_1 * S_1$	y^2	$S_{13} = S_3 * S_{12}$	$ay(y^2 - 2a - a)$
$S_7 = S_6 - S_4$	$y^2 - 2a$	$S_{14} = S_{10} - S_{13}$	$f(y)$

What was surprising that the participants found two solutions that has 11 and 13 steps! These solutions were awarded by additional points. The solution with 11 steps were found by Madalina Bolboceanu (Romania, Bitdefender) during the first round (Table 6). The solution with 13 steps were given by Henning Seidler and Katja Stumpp team (Germany, TU Berlin) during the second round. Both of the solution were based on the representation $f(y) = y((y^2 - 44)(y^2 - 66) - 22^2)$.

2.7 Problem "A promise"

2.7.1 Formulation

Young cryptographers, Alice, Bob and Carol, are interested in quantum computings and really want to buy a quantum computer. A millionaire gave them a certain amount of money (say, X_A

Table 6: **List of commands for the 11-step solution**

Command	Result	Command	Result
$S_1 = y$	y	$S_7 = S_6 - S_4$	$y^2 - 44 - 22$
$S_2 = S_1 * S_1$	y^2	$S_8 = S_6 * S_7$	$(y^2 - 44) * (y^2 - 44 - 22)$
$S_3 = 2$	2	$S_9 = S_4 * S_4$	22^2
$S_4 = 22$	22	$S_{10} = S_8 - S_9$	$(y^2 - 44) * (y^2 - 44 - 22) - 22^2$
$S_5 = S_3 * S_4$	44	$S_{11} = S_1 * S_{10}$	$f(y)$
$S_6 = S_2 - S_5$	$y^2 - 44$		

for Alice, X_B for Bob, and X_C for Carol). He also made them promise that they would not tell anyone, including each other, how much money everyone of them had received.

- Could you help the cryptographers to invent an algorithm how to find out (without breaking the promise) whether the total amount of money they have, $X_A + X_B + X_C$, is enough to buy a quantum computer?
- What weaknesses does your algorithm have (if someone breaks the promise)? Does it always protect the secret of the honest participants from the dishonest ones?

2.7.2 Solution

This problem is a particular case for the problem "A promise and money" for only three participants (see section 2.8).

2.8 Problem "A promise and money"

2.8.1 Formulation

A group of young cryptographers are interested in quantum computings and really want to buy a quantum computer. A millionaire gave them a certain amount of money (say, n cryptographers; X_i for each of them, $i = 1, \ldots, n$). He also made a promise from them that they would not tell anyone, including each other, how much money everyone of them had received.

- Could you help the cryptographers to invent an algorithm how to find out (without breaking the promise) whether the total amount of money they have, $\sum_{i=1}^{n} X_i$, is enough to buy a quantum computer?
- What do you think whether there are such algorithms protecting the secrets of honest participants from dishonest ones?
- What weaknesses does your algorithm have (if someone breaks the promise)? Does it always protect the secret of honest participants from dishonest ones?

2.8.2 Solution

Here we give an idea of the solution proposed by Mikhail Kudinov (Bauman Moscow State Technical University).

First of all, it is supposed that no one can buy a quantum computer himself without other participants. Let us assume that N' is the amount of money that one needs to buy a quantum computer and $N = nN'$, where n is the number of participants. The millionaire gave them X_i

money for $i \in \{1, \ldots, n\}$. Each of participants chooses random secrets $s_{i,j}$ uniformly so that

$$\sum_{j=1}^{n} s_{i,j} \equiv X_i \pmod{N}.$$

Then each of then gives the share $s_{i,j}$ to the owner of X_j by the secure channel. After this procedure, the owner of X_i has shares $s_{k,i}$ for each $k \in \{1, \ldots, n\}$. It is obvious that

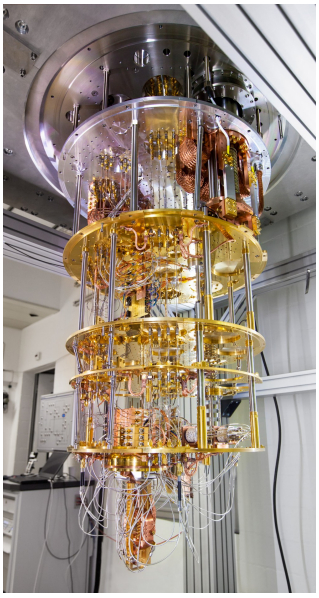
$$\sum_{j=1}^{n} \sum_{i=1}^{n} s_{i,j} = \sum_{i=1}^{n} X_i \pmod{N}.$$

Under the first suggestion, all participants can together calculate the common amount of money.

The main disadvantage of the algorithm, in addition to the suggestion, is a big amount of private communication (though the number of keys can be n for asymmetric schemes).

Analogically, many participants described algorithms similar to Schneier's calculating average salary algorithm [13]. In general, all such algorithms are vulnerable if $n - 1$ participants are dishonest. Some participants tried to describe a possibility to use a cryptosystem, that is homomorphic by "+" and preserves relation "<", as a general analysis.

The problem of the first school round is the same problem for $n = 3$ (score assignment was more loyal). Despite there was a quite big number of solutions for this problem in the student round, each solution had big or small lacks in analysis of the general case, in analysis of the algorithm advantages and disadvantages, in description of communications (number of private communications, what kind of cryptography is used, number of required private keys) and so on. As a result, there was no possibility to chose "best of the best" for 6 scores and we decided to give 5 scores as maximum. There were nine maximal-scored solutions.



IBM's 50 qubit quantum computing system [21]

2.9 Problem "16QAM"

2.9.1 Formulation

For sending messages, Alice and Bob use a fiber-optic communication via 16QAM technology. This technology allows to send messages whose alphabet consists of 16 letters, where each letter is usually encoded with a 4-bit Gray code. While a message is transmitted in the channel, single errors in codewords of the Gray code are possible.

Alice has read an interesting book and would like to share her enthusiasm with Bob! Alice sent a short fragment from the book to Bob. Due to the characteristics of the communication channel used, she divided the text into two parts and sent them separately. In the first part, she placed all of the 16 consonants that occurred in this fragment; in the second part, she placed vowels ("y" is a vowel), a space, a hyphen and punctuation marks. Then Alice also encoded the letters with Hamming code to be able to correct single errors. She applied a 7-bit Hamming code with the parity-check matrix whose columns are written in lexicographical order.

Bob received the following two parts of ciphertext (given in hexadecimal notation):

Part 1	Part 2
66674C36666F43D3C199900AA1AA325992A	66CA61967319CCD2CE76998CE6433332D19
67A59D9B4A8B69330D1BC000153367A5E33	B46784C65334E999A402ADA0265A99A6633
D30E6692D0F349D3321FFFF0ED706667A7F	33319B32D3299698CCC96986619967134CC
670D999679F4AA67561BA679B4AA54F34D5	B4CE23333334CC6730CE90170CCCD2CE669
AB0F4AACCF000055CE633670D9DA54CE37F	996A61999EA63332CCA4C3332D4CD3334CC
660DE19CD995335495523CCAAA8F1E03325	D3319994730CCCD3A6669D96A66999699B3
86CF48A98CD9B387FD9D546A99E9D200033	98640CC86CE619676AD4CD3308999866D33
3201513FE5B4AA00CCCE9667554CD2CCCB3	79321C33210B4C6732199B53218019A404C
330F32A666553CD756AC3E0674E9D369E1D	D2DE65A986663398CCCCCB5319CC6665997
C6A9999780007F00961E66465519FEA8B25	B96A63398CD9CCD2CD9A399A66339866619
14CCCB332AA63332CCCE6D2A99AACCCC004	98CD9CC325A6339CCE619998C04C66CE633
	996A61998CF66967334CC66CA6199865E$(0)_2$

Also, he received the following number sequence: 22, 19, 3, 3, 36, 53, 3, 33, 20, 28. Each number indicates how many consonants are contained between the punctuation marks.

Recover the text and find the main character of the book Alice has read!

2.9.2 Solution

Some details in the problem statement are insignificant. Namely, we could omit the step with the Gray code and mind that Alice substitutes 7-bit codewords of the Hamming code for each symbol in each part of the plaintext.

The crucial idea to broke the cipher Alice and Bob use is analyzing the frequency distribution in each part of the ciphertext. This helps to deduce the probable meaning of the most common symbols and form partial words. Tentative search for combinations of consonants and vowels giving actual words in English expands the partial solution. Frequencies of pairs of letters also give an improvement but it could seem inessential. At last, one can employ search engine on the Internet to find the fragment of the book that Alice sent to Bob.

Let us consider a possible solution. Alice uses the Hamming code with the parity check matrix H and the corresponding generator matrix G, where

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

First, rewrite each part of the given ciphertext in the binary form. Split them into 7-bit words and correct errors using the parity check matrix H. One can decode the Hamming code into a 4-bit Gray code but it is not a necessary step for the solution. Calculating frequencies of codewords separately in each part of the given ciphertext, we put them in Table 7.

Table 7: **Frequencies of Hamming codewords in the text**

Gray code	Hamming code	Frequency
1011	0110011	46
0010	0101010	30
1001	0011001	24
0001	1101001	24
0011	1000011	19
0000	0000000	15
0110	1100110	13
1100	0111100	8
1111	1111111	8
1101	1010101	7
0100	1001100	6
1110	0010110	5
1010	1011010	5
0101	0100101	4
1000	1110000	4
0111	0001111	2

a) Part 1

Gray code	Hamming code	Frequency
0100	1001100	85
1011	0110011	50
1001	0011001	33
0001	1101001	26
1010	1011010	17
0011	1000011	9
0000	0000000	8
1110	0010110	7
1100	0111100	2
0010	0101010	1
1000	1110000	1
0111	0001111	0
0101	0100101	0
1101	1010101	0
0110	1100110	0
1111	1111111	0

b) Part 2

Compare the frequencies obtained with those of letters in the English language. The suitable frequency distribution can be found in [12], which is cited, e. g., at [20]. According to Lewand, arranged from most to least common in appearance, the letters are:

`e t a o i n s h r d l c u m w f g y p b v k j x q z`.

We start with vowels, punctuation marks, spaces, and a hyphen, which are placed in Part 2. Make a guess that the most frequent symbol in Part 2 is the space. It is also worth to note that most of punctuation marks are followed by a space in contrast to a hyphen, which is usually embraced by letters. Using letter frequencies, we determine the probable spaces, vowels, and hyphen, and construct the following partial solution for this part of the plaintext (the sign `#` substitutes punctuation):

```
ee ae e oe o e ua iaia# e oo oy-oy i o ea ee# u# ea# auae o ie ea o e aoy a oe
o i a i eae# a i o o o eae a oo o i o iee ay ue aeii o aa aie# uuay# e uai uy oy
oe i a e ea i e eae# i e ee oeee o e a a ee a# e e a uy ee e i a e oe o ee a a#
```

Let us turn to Part 1, which contains 16 consonants occurring in the fragment of the book. Let us order the codewords of the Hamming code from most to least frequent in Part 1, as it is shown in Table 7a. Denote the 7-bit codewords by hexadecimal numbers from `0` till `F`. Then we get the following ciphertext of 220 symbols in length that is splitted into 10 pieces (according to the number sequence given in the task):

```
023402C43E0251412B0103  02C1B32407551003703  4A3  B46  33A4884CE02E804020631094106311739943
1675510A0040C1068047266101D10619FF56D4031A00048090103  355
025108B315023021A3020246102173994  E2333C72410275585D46  021281BD102021A0202631016055
```

Then we match symbol frequencies in Part 1 of the ciphertext with those of consonants in the English alphabet. The first five pairs are like as follows: `0 - t`, `1 - n`, `2 - s/h`, `3 - s/h`, `4 - r`.

The bigram `th` is the most frequent in English. This allows us to make a suggestion that `2` substitutes `h` and `3` substitutes `s`. Then we obtain a partial solution for Part 1 and, combining with one for Part 2, get the following pieces of the plaintext given in Table 8. It is not difficult to recognize words `these are the` at the beginning in (1). Also, we can see `the` as the first word in (2) and (8).

Table 8: **Partial plaintext**

No.	Partial plaintext
(1)	`thsrthCrsEth5nrnhBtnts` `ee ae e oe o e ua iaia#`
(2)	`thCnBshrt755ntts7ts` `e oo oy-oy i o ea ee#`
(3)	`rAs` `u#`
(4)	`Br6` `ea#`
(5)	`ssAr88rCEthE8trtht6snt9rnt6snn7s99rs` `auae o ie ea o e aoy a oe o i a i eae#`
(6)	`n6755ntAttrtCnt68tr7h66ntnDnt6n9FF56DrtsnAtttr8t9tnts` `a i o o o eae a oo o i o iee ay ue aeii o aa aie#`
(7)	`s55` `uuay#`
(8)	`th5nt8Bsn5thsthnAsththr6nthn7s99r` `e uai uy oy oe i a e ea i e eae#`
(9)	`EhsssC7hrnth75585Dr6` `i e ee oeee o e a a ee a#`
(10)	`thnh8nBDnththnAthth6sntn6t55` `e e a uy ee e i a e oe o ee a a#`

The best idea for the next step is to search through the English dictionary for words that have given vowels in the prescribed order. It is possible to use one of the tools for pattern recognition available on the Internet, e. g., [19]. Advanced participants of the Olympiad implemented some computer programs on their own.

Consider several examples. We have a word with consonants `s55` and vowels `uuay` in (7), and the last two consonants are identical. The only match is `usually`, so we assume that `5` substitutes the letter `l`. The pattern `auae` in combination with double `s` gives us two possibilities in (5) – `assuage` and `sausage`. In any case, it seems like `A` means `g`. Then we have `rugs` in (3). The pattern `uai` and consonants `5nt8B` lead us to `lunatic` in (8), so `8` probably means `c`.

At this point we revise our matching the letters and their frequencies corresponding to the Part 1 of the ciphertext. Let us look at the first eight letters with large frequencies: `t n h s r l 6 7/c`. We can see that the letter `d` has still been hidden. According to the Lewand distribution it is the most probable that `6` means `d`. Then (4) contains `Brd` and `ea` what gives us possible words `beard` and `bread`. Therefore, it seems like `B` substitutes `b`.

A thoroughly analysis of the remaining ciphertext and search for words by patterns and number of letters eventually lead us to the plaintext (with punctuation replaced by `#`):

```
these are the mores of the lunar inhabitants# the moon boy-shorty will not eat
sweets# rugs# bread# sausage or ice cream of the factory that does not print
ads in newspapers# and will not go to treatment a doctor who did not invented
any puzzle advertising to attract patients# usually# the lunatic buys only
those things that he read in the newspaper# if he sees somewhere on the wall
a clever ad# then he can buy even the thing that he does not need at all#
```

This is a fragment of the fairytail novel “Dunno on the Moon” by Russian writer Nikolay Nosov. The title character of the novel is a boy-shorty Dunno. The problem was completely solved by 13 teams in the second round and by Samuel Tang (Hong Kong, Black Bauhinia) in the first round.

The best solutions were proposed by the team of Irina Slonkina, Mikhail Sorokin, and Vladimir Bobrov (Bauman Moscow State Technical University), and the team of Vladimir Paprotski, Dmitry Zarembo, and Karina Kruglik (Belarusian State University).

2.10 Problem "APN + Involutions"

The first three questions **Q1, Q2, Q3** were given as the problem "APN + Involutions" in the first round. The extended version of the task for the second round included also the question **Q4** that contains open problems.

2.10.1 Formulation

Alice wants to construct a block cipher with heavy use of **involutions** as subcomponents; this minimizes difference between the algorithms for encryption and decryption. She knows that **APN permutations** are the best choice of subcomponents to resist attacks based on differential technique. She wants to construct a set of APN permutations that are involutions for every $n \geqslant 2$.

Alice knows that any involution can be expressed as the product of disjoint **transpositions**. So, she decides to study the following involution

$$g = \prod_{i=1}^{d} (\alpha_i, {\alpha'}_i),$$

where $\{\alpha_i, {\alpha'}_i\} \cap \{\alpha_j, {\alpha'}_j\} = \emptyset$ for all $i, j \in \{1, ..., d\}$, $i \neq j$, $1 \leqslant d \leqslant 2^{n-1}$.

Alice needs your help to get APN permutations among such involutions g. Find answers to the following questions!

Q1 Let

$$\Lambda(g) = \big\{\alpha_i \oplus {\alpha'}_i : \ i = 1, ..., d\big\}, \quad \widehat{\Lambda}(g) = \big[\alpha_i \oplus {\alpha'}_i : \ i = 1, ..., d\big],$$
$$\mathrm{B}(g) = \big\{x \oplus y : \ \{x, y\} \subseteq \mathrm{FixP}(g), \ x \neq y\big\}, \quad \widehat{\mathrm{B}}(g) = \big[x \oplus y : \ \{x, y\} \subseteq \mathrm{FixP}(g), \ x \neq y\big],$$

where $\mathrm{FixP}(g)$ is the set of all **fixed points** of g, i. e. $\mathrm{FixP}(g) = \{x \in \mathbb{F}_2^n : \ g(x) = x\}$.

Suppose that g is an APN permutation. Get necessary conditions for multisets $\widehat{\Lambda}(g)$, $\widehat{\mathrm{B}}(g)$ and sets $\Lambda(g)$, $\mathrm{B}(g)$. Prove that if your conditions do not hold, then g is not an APN permutation.

Q2 Let

$$\mathrm{d}_{a,b}(g) = |\{x \in \mathbb{F}_2^n : g(x \oplus a) \oplus g(x) = b\}|, \quad a, b \in \mathbb{F}_2^n.$$

Let g be an involution and APN. Find $\mathrm{d}_{a,a}(g)$ for each nonzero $a \in \mathbb{F}_2^n$.

Q3 Can you get the nontrivial upper bound on $|\mathrm{FixP}(g)|$?

Q4 Let M_n be the set of all n-bit involutions that are APN permutations.

(a) Can you find the cardinality of M_n for $n = 2, 3, 4$?

(b) Can you find the cardinality of M_n for $n = 5$?

(c) **Bonus problem (extra scores, a special prize!)**
Let $n \geqslant 6$. Can you get the lower and the upper bounds for the cardinality of M_n? Can you describe involutions from M_n? Can you suggest constructions for involutions from M_n?

Note that the mapping $x \mapsto x^{-1}$ in the Galois field $GF(2^n)$ belongs to M_n for odd $n \geqslant 3$.

Remark. Let us recall relevant definitions.

- $\mathbb{F}_2^n$ is the vector space of dimension over $\mathbb{F}_2 = \{0, 1\}$.
- A vector $x \in \mathbb{F}_2^n$ has the form $x = (x_1, ..., x_n)$, where $x_i \in \mathbb{F}_2$. For two vectors $x, y \in \mathbb{F}_2^n$ their sum is $x \oplus y = (x_1 \oplus y_1, ..., x_n \oplus y_n)$, where $\oplus$ stands for XOR operation.
- Let $\widehat{X} = \left[x_1, ..., x_d\right]$ be a multiset with the underlying set $\mathbb{F}_2^n$, where $x_1, ..., x_d \in \mathbb{F}_2^n$. Note that all elements in a set are distinct. Unlike a set, a multiset allows for multiple instances for each of its elements.
- A **permutation** s is a mapping from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ such that $s(x) \neq s(y)$ for all $x, y \in \mathbb{F}_2^n$, $x \neq y$.
- An **involution** s is a permutation that is its own inverse, $s^2(x) = s(s(x)) = x$ for all $x \in \mathbb{F}_2^n$.
- For any different vectors $\alpha, \beta \in \mathbb{F}_2^n$, a permutation s is called a **transposition** if $s(\alpha) = \beta$, $s(\beta) = \alpha$ and $s(x) = x$ for all $x \in \mathbb{F}_2^n \backslash \{\alpha, \beta\}$; it is denoted by $s = (\alpha, \beta)$.
- A permutation s is called **APN** (Almost Perfect Nonlinear) if, for every nonzero $a \in \mathbb{F}_2^n$ and every $b \in \mathbb{F}_2^n$, the equation $s(x \oplus a) \oplus s(x) = b$ has at most 2 solutions.

2.10.2 Solution

Q1 Let $a \in \Lambda(g)$. Hence, $a = x \oplus y$, where $y = g(x)$ and $(x, y) = (\alpha_i, \alpha'_i)$ for some i. Then

$$g(x \oplus a) = g(y) = x = y \oplus a = g(x) \oplus a.$$

Let $a \in \mathrm{B}(g)$. Hence, $a = x \oplus y$, where $x, y \in \mathrm{FixP}(g)$. Then

$$g(x \oplus a) = g(y) = y = x \oplus a = g(x) \oplus a.$$

Thus, $\mathrm{d}_{a,a}(g) \geqslant 2$ for any vector $a \in \Lambda(g) \cup \mathrm{B}(g)$.

Let g be an APN permutation. Then $\mathrm{d}_{a,a}(g) = 2$. Hence, the multiplicity of all elements from $\Lambda(g)$ and $\mathrm{B}(g)$ is 1. Thus, $\Lambda(g) = \widehat{\Lambda}(g)$ and $\mathrm{B}(g) = \widehat{\mathrm{B}}(g)$. Note that $\Lambda(g) \cap \mathrm{B}(g) = \emptyset$.

Q2 Since g is an APN permutation, then $\mathrm{d}_{a,a}(g) \leqslant 2$. As we get in **Q1**, $\mathrm{d}_{a,a}(g) = 2$ for any vector $a \in \Lambda(g) \cup \mathrm{B}(g)$. Let us prove that $\mathrm{d}_{a,a}(g) = 0$ for $a \notin \Lambda(g) \cup \mathrm{B}(g)$.

Let a be a nonzero vector and x be a solution of $g(x \oplus a) \oplus g(x) = a$. Since g is a permutation, then either $x \in \mathrm{FixP}(g)$ or $x = \alpha_i$ $(x = \alpha'_i)$ for some i. Consider two cases:

1. Let $x \in \mathrm{FixP}(g)$. Then, $g(x \oplus a) \oplus g(x) = a$ implies $g(x \oplus a) = x \oplus a$. Hence, $x \oplus a \in$ FixP(g). As a result, $a \in B(g)$.
2. Without loss of generality, let $x = \alpha_i$ for some i and $y = x \oplus a$. If $y \in \mathrm{FixP}(g)$, then $g(x \oplus a) \oplus g(x) = a$ implies $g(x) = x$, which is a contradiction. Hence, without loss of generality, $y = \alpha'_j$ for some j (so, we have $\alpha_i \oplus \alpha'_j = a$). Then

$$g(\alpha_i \oplus a) \oplus g(\alpha_i) = a \;\Rightarrow\; g(\alpha'_j) \oplus \alpha'_i = a \;\Rightarrow\; \alpha_j \oplus \alpha'_i = a.$$

Let us show that α'_i and α_j is also solutions. Indeed,

$$g(\alpha'_i \oplus a) \oplus g(\alpha'_i) = g(\alpha_j) \oplus \alpha_i = \alpha'_j \oplus \alpha_i = a$$

and

$$g(\alpha_j \oplus a) \oplus g(\alpha_j) = g(\alpha'_i) \oplus \alpha'_j = \alpha_i \oplus \alpha'_j = a.$$

Thus, if $i \neq j$, we get at least 3 solutions that is contradiction for the APN property of g. Hence, $j = i$ and $a \in \Lambda(g)$.

Q3 Let us prove that $|\mathrm{FixP}(g)| \leqslant 1 + (2^{n-1}-1)^{1/2}$.

The involution g is APN. From **Q1** we have

$$\mathrm{B}(g) \cap \Lambda(g) = \emptyset. \tag{1}$$

Let $q = |\mathrm{FixP}(g)|$. Since g is an involution, we have that q is even. From equality (1) and $\Lambda(g) \cup \mathrm{B}(g) \subseteq \mathbb{F}_2^n \backslash \{\mathbf{0}\}$ it follows that

$$|\Lambda(g)| + |\mathrm{BB}(g)| \leqslant 2^n - 1. \tag{2}$$

Since $|\mathrm{B}(g)| = \binom{q}{2}$, $|\Lambda(g)| = 2^{n-1} - q/2$, we have

$$|\Lambda(g)| + |\mathrm{BB}(g)| = q(q-1)/2 + 2^{n-1} - q/2.$$

From inequality (2), we get

$$q(q-1)/2 + 2^n - q \leqslant 2^n - 1.$$

Thus,

$$q(q-2)/2 \leqslant 2^{n-1} - 1,$$

i. e.

$$q \leqslant 1 + (2^{n-1}-1)^{1/2}.$$

Q4 (a) It could be computationally verified that $M_2 = \emptyset$ and $|M_3| = 224$. Then, it is known [3] that there are no APN permutations for $n = 4$. Hence, $M_4 = \emptyset$.

(b) Let us recall several definitions. A function $A : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is *affine* if $A(x \oplus y) = A(x) \oplus A(y) \oplus A(\mathbf{0})$ for any $x, y \in \mathbb{F}_2^n$. Two functions $F, G : \mathbb{F}_2^n \to \mathbb{F}_2^n$ are called *affine equivalent* if there exist affine permutations A_1, A_2 such that $F = A_1 \circ F \circ A_2$. It is easy to see that the APN permutation property of a function is an invariant under the affine equivalence. There exist [3] only five the affine equivalence classes of APN permutations. Moreover, by [3, theorem 3] only one class contains functions together with their inverses. Hence, only this class of APN permutations can contain involutions. The representative of this class is the famous inverse function over the finite field: $F(x) = x^{-1}$ for nonzero x and $F(0) = 0$ (here, functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ are considered as functions over the finite field of order 2^n). The inverse function is an involution. Thus, all APN involutions for $n = 5$ are affine equivalent to the inverse function.

(c) There were no interesting suggestions by the participants for these open problems.

The unique full correct solution in the first round was proposed by Henning Seidler (Germany, TU Berlin). In the second round, the best solution for 11 scores was proposed by the team of Kristina Geut, Sergey Titov, and Dmitry Ananichev (Russia, Ural State University of Railway Transport, Ural Federal University).

2.11 Problem "Sharing"

2.11.1 Formulation

Bob is interested in studying mathematical countermeasures to side-channel attacks on block ciphers. He found out that techniques such as special sharings of functions can be applied. Now he is thinking about the following mathematical problem in this approach.

Let $\mathcal{F}$ denote the set of **invertible functions** (**permutations**) from $\mathbb{F}_2^4$ to $\mathbb{F}_2^4$ and $\mathcal{F}^n$ denote the set of invertible functions from $(\mathbb{F}_2^4)^n$ to $(\mathbb{F}_2^4)^n$. Let $F \in \mathcal{F}^n$ be

$$F(x_1, x_2, \ldots, x_n) = (F_1(x_1, x_2, \ldots, x_n), F_2(x_1, x_2, \ldots, x_n) \ldots, F_n(x_1, x_2, \ldots, x_n)),$$

with component functions $F_i : (\mathbb{F}_2^4)^n \to \mathbb{F}_2^4$, $i = 1, \ldots, n$.

For any $f \in \mathcal{F}$, a function $F \in \mathcal{F}^n$ is called a **sharing** of f if

$$\sum_{i=1}^{n} F_i(x_1, x_2, \ldots, x_n) = f\left(\sum_{i=1}^{n} x_i\right) \quad \text{for all} \quad (x_1, x_2, \ldots, x_n) \in (\mathbb{F}_2^4)^n.$$

Moreover, F is a **non-complete** sharing of f if F is a sharing of f with the additional property that each component function F_i is independent of x_i.

Bob needs your help to study functions for which non-complete sharing exists. Find answers to the following questions!

Q1 Let $\mathcal{A}$ denote the set of **affine functions** from $\mathbb{F}_2^4$ to $\mathbb{F}_2^4$. Two functions $f, g \in \mathcal{F}$ are **affine equivalent** if there exist $a, b \in \mathcal{A}$ such that $g = b \circ f \circ a$.

Let f, g be two functions in the same affine equivalence class of $\mathcal{F}$ and let F be a non-complete sharing of f. Derive from F a non-complete sharing for g.

All functions of the same affine equivalence class have the same degree. It is known [4] that this equivalence relation partitions $\mathcal{F}$ into 302 classes: 1 class corresponds to $\mathcal{A}$, 6 classes contain quadratic functions, 295 classes contain cubic functions.

Also, Bob knows that when $n \geqslant 5$, there exists a non-complete sharing for each $f \in \mathcal{F}$ (it can be shown by construction). When $n = 2$ a non-complete sharing exists only for the functions in $\mathcal{A}$. When $n = 3$, non-complete sharings exist for $\mathcal{A}$ and also for 5 out of the 6 equivalence classes containing quadratic functions. When $n = 4$, non-complete sharings exist for $\mathcal{A}$, for all 6 quadratic equivalence classes and for 5 cubic classes.

Q2 Bonus problem (extra scores, a special prize!)

Find a concise mathematical property that a function $f \in \mathcal{F}$ must have in order that a non-complete sharing F exists for $n = 3, 4$.

Q3 Bonus problem (extra scores, a special prize!)

Generalize to functions over $\mathbb{F}_2^5$, $\mathbb{F}_2^6$.

2.11.2 Solution

Q1 Let f, g be two functions in the same affine equivalence class of $\mathcal{F}$, that is $g = b \circ f \circ a$ for some $a, b \in \mathcal{A}$, and let $F \in \mathcal{F}^n$ be a non-complete sharing of f. At first, one can notice that since f, g are invertible, the mappings a, b must be invertible as well. Let us denote

$$a(x) = Ax + a', \quad x \in \mathbb{F}_2^4,$$

$$b(x) = Bx + b', \quad x \in \mathbb{F}_2^4,$$

where A, B are nonsingular binary matrices of order 4×4 and $a', b' \in \mathbb{F}_2^4$.

Using components functions $\{F_i\}_{i=1}^n$ of F, we define the invertible function $G \in \mathcal{F}^n$ with components functions

$$G_j(x_1, x_2, ..., x_n) = \begin{cases} BF_1(Ax_1 + a', Ax_2, ..., Ax_n) + b', & j = 1, \\ BF_j(Ax_1 + a', Ax_2, ..., Ax_n), & j \neq 1, \end{cases}$$

where $j = 1, 2, ..., n$.

Then for any $(x_1, x_2, ..., x_n) \in \left(\mathbb{F}_2^4\right)^n$, it holds

$$\sum_{j=1}^n G_j(x_1, x_2, ..., x_n) = BF_1\left(Ax_1 + a', Ax_2, ..., Ax_n\right) + b' +$$

$$+ \sum_{j=2}^n BF_j\left(Ax_1 + a', Ax_2, ..., Ax_n\right) = B\left(\sum_{j=1}^n F_j\left(Ax_1 + a', Ax_2, ..., Ax_n\right)\right) + b' =$$

$$= Bf\left(Ax_1 + a' + Ax_2 + \ldots + Ax_n\right) + b' = Bf\left[A\left(\sum_{i=1}^n x_i\right) + a'\right] + b' =$$

$$= b \circ f \circ a\left(\sum_{i=1}^n x_i\right) = g\left(\sum_{i=1}^n x_i\right).$$

Therefore, the function $G \in \mathcal{F}^n$ defined as

$$G(x_1, x_2, ..., x_n) = (G_1(x_1, x_2, ..., x_n), G_2(x_1, x_2, ..., x_n), ..., G_n(x_1, x_2, ..., x_n)),$$

is a sharing of g.

From non-completeness of F it follows that G_j, which is in fact an affine transformation of F_j, does not depend on x_j. Hence, G is a non-complete sharing of g.

Q2-Q3 These open problems were not solved completely during the Olympiad. Nevertheless, one perspective solution was proposed by the team of Victoria Vlasova, Mikhail Polyakov, and Alexey Chilikov (Bauman Moscow State Technical University). They found a sufficient condition for the existence of non-complete sharing for $n = 3$. Let us describe it here.

Let $\mathrm{wt}(y)$ be the Hamming weight of a binary vector y. For $\sigma \in \mathbb{F}_2$, we denote

$$\delta_\sigma(y) = \begin{cases} y, & \sigma = 1, \\ \mathbf{0}, & \sigma = 0, \end{cases}$$

where $\mathbf{0}$ is a zero vector of the same dimension as y.

Let V be a vector space over the field K and assume that for the invertible function $f : V \to V$ it holds

$$\sum_{\sigma \in \mathbb{F}_2^n} (-1)^{\mathrm{wt}(\sigma)} f\left(\sum_{i=1}^n \delta_{\sigma_i}(x_i)\right) = 0, \tag{3}$$

then there exists a non-complete sharing for f. Further we conider the case $n = 3$.

Indeed, for any $(x_1, x_2, x_3) \in V^3$ put

$$F_1(x_1, x_2, x_3) = f(x_2) - f(x_2 + x_3),$$

$$F_2(x_1, x_2, x_3) = f(x_3) - f(x_1 + x_3),$$
$$F_3(x_1, x_2, x_3) = f(x_1) - f(x_1 + x_2).$$

It is clear that every $F_i : V^3 \to V$ does not depend on x_i, where $i = 1, 2, 3$. Consider the expression

$$\sum_{i=1}^{3} F_i(x_1, x_2, x_3) = f(x_2) - f(x_2 + x_3) + f(x_3) - f(x_3 + x_1) + f(x_1) - f(x_1 + x_2) =$$
$$= \sum_{\sigma \in \mathbb{F}_2^3} (-1)^{\mathrm{wt}(\sigma)} f\left(\sum_{i=1}^{3} \delta_{\sigma_i}(x_i)\right) + f(x_1 + x_2 + x_3) - f(0) = f(x_1 + x_2 + x_3) - f(0).$$

Without loss of generality we assume that $f(0) = 0$, otherwise we can consider the initial problem for the function $g(x) = f(x) - f(0)$ with $g(0) = 0$ and which, by the arguments from **Q1**, has non-complete sharing if and only if f does.

Finally

$$\sum_{i=1}^{3} F_i(x_1, x_2, x_3) = f(x_1 + x_2 + x_3),$$

that concludes the proof.

It was also shown by the authors that the condition (3) is necessary for the existence of non-complete sharing of f for any n.

Taking $V = \mathbb{F}_2^m$ with $m = 4, 5, 6$ and $K = \mathbb{F}_2$ one can obtain a solution of **Q2, Q3** for the case $n = 3$.

2.12 Problem "Factoring in 2019"

2.12.1 Formulation

Nicole is learning about the RSA cryptosystem. She has chosen random 500-bit prime numbers p and q, $2^{499} \leqslant p, q < 2^{500}$, and computed $n = p \cdot q$. Being a curious and creative person, she has also combined the three numbers in funny ways. Her favorite one is an integer h such that

$$h \equiv 3^{2019} p^2 + 5^{2019} q^2 \pmod{n^2 + 8 \cdot 2019}.$$

Unfortunately, she has lost the paper where she wrote the two prime numbers. Luckily, she remembers n and h. Help Nicole to recover p and q.

$$\begin{aligned} n \;=\; & 40763613025504836845249840044831561583564626405535158138667037 \\ & 18791672670905308860844304055285019651507728831663677166092475 \\ & 16155419756121537288444995708421977847213953345126368990185271 \\ & 10259760189356588305406519080647582874212687596214191915933827 \\ & 67252094717222418132289251314647500491996323400002019, \end{aligned}$$

$$
\begin{aligned}
h \;=\; & 7830799927833657758696152811024002692382891492752691194950119 6 \\
& 6454949775637356998539355466113271719836871709311181256664903 1 \\
& 1734281844963358864709854461215127803513145423478665313650088 7 \\
& 0883047099654288891241821353207362290372720539680784860373583 5 \\
& 7265363088368590691670158736223664912689571965666329382550122 3 \\
& 9708879962925260124942806243225473893576430461028161326422564 1 \\
& 7499027286468001256009599212578383223023458925765092934836426 8 \\
& 4811749406546352920185960074752189295725810403319544101402343 2 \\
& 3658152920139218532763567492345929074924183159066190396513251 4 \\
& 2154451518308886658505820006667836934411881.
\end{aligned}
$$

2.12.2 Solution

This problem is based on a (simplified) variation of the Coppersmith method.

Let $m = n^2 + 8 \cdot 2019$. It is a composite number with unknown factors. The idea is to find an integer a such that numbers

$$
\begin{aligned}
a_1 &= a \cdot 3^{2019} \mod m, \text{ and} \\
a_2 &= a \cdot 5^{2019} \mod m
\end{aligned}
$$

are small enough and $a_1 p^2 + a_2 q^2$ exceeds the modulus m by a small amount and can be recovered from $a \cdot h \mod m$. This can be done using the Lagrange-Gauss algorithm (which is a special case and the building block of the LLL algorithm). Let Λ be the lattice spanned by the two vectors

$$
\begin{aligned}
v_1 &= \big(1, \;\; (5^{2019} \cdot (3^{2019})^{-1} \mod m)\big), \\
v_2 &= \big(0, \;\; m\big).
\end{aligned}
$$

Consider an arbitrary vector $v = (a_1, a_2)$ in this lattice. It is easy to verify that

$$
a_1 p^2 + a_2 q^2 \equiv a_1 \cdot h \cdot (3^{2019})^{-1} \pmod m.
$$

The lattice reduction guarantees to find such vector v with the norm

$$
\|v\| = \sqrt{a_1^2 + a_2^2} \le 2^{(d-1)/4} (\det \Lambda)^{1/d} = \sqrt{m} / \sqrt[4]{2},
$$

where $d = 2$ is the dimension of the lattice. In particular,

$$
|a_1 p^2 + a_2 q^2| \le n(p^2 + q^2) < n(p+q)^2 < 10n^2,
$$

where the last two inequalities follow from balancedness of the primes (i.e., $\max(p, q) \le 2 \min(p, q)$).

It follows that there exists an integer $z, |z| < 10$, such that

$$
a_1 \cdot h \cdot (3^{2019})^{-1} \mod m + zm = a_1 p^2 + a_2 q^2.
$$

As a result, we obtain an equation in p^2 and q^2. By replacing $p = n/q$, we obtain a biquadratic equation in q which is easy to solve and factor n.

The final solution is:

$$\begin{aligned} p &= 201900007587815418168112981041447702234681820917519452487920889092150114454704800795372227128569035026411608157924118958739320260266419989959402141438 3, \\ q &= 201900007397349419452133980568209395918226574608399559482639375363166928917582785166666801416711943938654328985094073488580682612071817972924264102689 3. \end{aligned}$$

The best solution was proposed by Alexey Zelenetskiy, Mikhail Kudinov, and Denis Nabokov team (Russia, Bauman Moscow State Technical University).

2.13 Problem "TwinPeaks3" (online)

2.13.1 Formulation

As Bob's previous cipher `TwinPeaks2` (NSUCRYPTO-2018) was broken again, he finally decided to read some books on cryptography. His new cipher is now inspired by practical ciphers, while the number of rounds was reduced a bit for better performance.

Not only the best techniques were adopted by Bob, but also he decided to enhance his cipher by security through obscurity, so the round functions are now unknown. The only thing known about these functions is that they are the same for odd and even rounds.

New Bob's cipher works as follows. A message X is represented as a binary word of length 128. It is divided into four 32-bit words a, b, c, d and then the following round transformation is applied 32 times:

$$(a, b, c, d) \leftarrow (b, c, d, a \oplus (F_i(b, c, d)))$$
$$F_i = F_1 \text{ for odd rounds and } F_i = F_2 \text{ for the rest.}$$

Here F_1 and F_2 are secret functions accepting three 32-bit words and returning one word; and $\oplus$ is the binary bitwise XOR. The concatenation of the final a, b, c, d is the resulting ciphertext Y for the message X.

Agent Cooper again wants to read Bob's messages. He caught the ciphertext

$$Y = \texttt{e473f19a247429ab33b66268d57dd241}$$

(the ciphertext is given in hexadecimal notation, the first byte is `e4`).

He was also able to gain access to Bob's testing server with encryption and decryption routines, using the secret key. Here it is [17]. Unfortunately, the version of software available on this server is not final. So, the decryption routine is incomplete and only uses keys in the reverse order, which is not sufficient for decryption:

$$(a, b, c, d) \leftarrow (b, c, d, a \oplus (F_i(b, c, d)))$$
$$F_i = F_2 \text{ for odd rounds and } F_i = F_1 \text{ for the rest.}$$

The server can also process multiple blocks of text at a time: they will be processed one-by-one and then concatenated, as in the regular ECB cipher mode of operation. Ciphertexts and plaintexts are given and processed by the server in hexadecimal notation.

Help Cooper to decrypt Y.

2.13.2 Solution

Let f_i be the round transformation of round i:

$$f_i : \ (a, b, c, d) \leftarrow (b, c, d, a \oplus (F_{k(i)}(b, c, d))),$$

where $k(i) = 1$ for odd i and $k(i) = 2$ for the rest.

Hence, we can represent the encryption transformation E as

$$E = (f_1 f_2)^{16}.$$

Let I be the incomplete decryption transformation described in the problem statement. The encryption and the incomplete decryption processes only differ in key order, so I can be written in terms of f_i:

$$I = (f_2 f_1)^{16}.$$

The decryption transformation E^{-1} can be represented as

$$E^{-1} = (f_2^{-1} f_1^{-1})^{16},$$

where f_i^{-1} is the inverse of f_i and is given by the following transformation:

$$f_i^{-1} : \ (a, b, c, d) \leftarrow (d \oplus (F_{k(i)}(a, b, c)), a, b, c)$$

Thus, to apply E^{-1} to the ciphertext one should be able to compute $F_1(x, y, z)$ and $F_2(x, y, z)$ that are secret. To recover these functions a *slide attack* can be used.

The idea is to find words $x = (x_1, x_2, x_3, x_4)$ and $y = (y_1, y_2, y_3, y_4)$ such that $f_i(x) = y$. If such a pair is found, then F_i can be found as

$$F_i(x_2, x_3, x_4) = y_4 \oplus x_1.$$

We use the following idea to find a desired pair: if $Ef_i(x) = E(y)$, then $f_i(x) = y$. Let us start with F_1. We need a pair of x and y such that $Ef_1(x) = E(y)$. This relation can be written as

$$(f_1 f_2)^{16} f_1(x) = (f_1 f_2)^{16}(y)$$
$$f_1 (f_2 f_1)^{16}(x) = (f_1 f_2)^{16}(y)$$
$$f_1 I(x) = E(y)$$

We come to a conclusion that if $f_1 I(x) = E(y)$, then $f_1(x) = y$. The condition $f_1 I(x) = Ey$ can be checked by using the definition of f_1: if $(I(x))_2 = (E(y))_1$, $(I(x))_3 = (E(y))_2$ and $(I(x))_4 = (E(y))_3$, then it is *likely* that $f_1 I(x) = E(y)$. The probability of false positives is approximately 2^{-96} for random F_i functions. So, it can be considered as negligible. Both $I(x)$ and $E(y)$ are available on the encryption oracle for arbitrary x and y as the incomplete decryption and the encryption routines respectively.

To find $F_i(a, b, c)$, let us brute force over x and y of the following forms: $x = (X, a, b, c)$ and $y = (a, b, c, X')$. According to the birthday paradox, a desired pair can be found in $2 * 2^{16}$ operations average (instead of 2^{32} if we lock X or X' to some constant value).

As soon as we find such a pair x and y, we can compute $F_1(a, b, c)$ and apply f_1^{-1} to the ciphertext and decrypt the last round. Then F_2 can be found the same way by replacing I and E with each other due to the symmetry. By doing this round by round, we decrypt the whole ciphertext and get the desired message (in hexadecimal notation)

`acherrypieplease`

The reference implementation of this attack requires 2^{22} blocks of text to be encrypted and 10 minutes of time average. It is important to use the server's ability to process multiple blocks of text at a time to minimize the amount of HTTP requests.

Four teams successfully solved the problem using the same method.

2.14 Problem "Curl27"

2.14.1 Formulation

Bob is developing the 3OTA infrastructure and has designed a new hash function Curl27 for it. A distinguishing feature of the infrastructure is the ternary logic: trits from the set $\mathbf{T} = \{0, 1, -1\}$ are used instead of bits, ternary strings and words are used instead of binary ones. The Curl27 hash function is defined below. Its implementation in Java can be found in [18].

Find a collision for Curl27, that is, different ternary strings X and X' such that Curl27$(X) =$ Curl27(X'). Submit colliding strings as two lines of trits separated by commas. An example of a (wrong!) solution is:

```
-1,1,0,1,1,0
-1,-1,1,0,1,1,-1,0
```

Description of Curl27. The Curl27 function maps a ternary string X of arbitrary length to a hash value from $\mathbf{T}^{243}$. When hashing, an auxiliary sponge function Curl27-f: $\mathbf{T}^{729} \to \mathbf{T}^{729}$ is used. The hashing algorithm:

1. Pad X with zeros to make its length a multiple of 243. Divide the resulting string into blocks $X_1, X_2, \ldots, X_d \in \mathbf{T}^{243}$.

2. Prepare the state $W = W_0 W_1 W_2 \in \mathbf{T}^{729}$ consisting of words $W_i \in \mathbf{T}^{243}$. Initialize the state by filling W_0 and W_2 with zeros and W_1 with the encoded initial (before padding) length of X. The length is encoded by a ternary word according to the little-endian conventions: less significant trits go first. For example, the length $25 = 1 - 3^1 + 3^3$ is presented by the word $\underbrace{1\bar{1}01000\ldots0}_{243}$. Here $\bar{1}$ stands for -1.

3. For $i = 1, 2, \ldots, d$, do: $W_0 \leftarrow X_i$, $W \leftarrow$ Curl27-f(W).

4. Return W_0.

Description of Curl27-f. In Curl27-f the S-box

$$S\colon \mathbf{T}^3 \to \mathbf{T}^3, \quad (a,b,c) \mapsto (F(a,b,c), F(b,c,a), F(c,a,b))$$

is used. Here

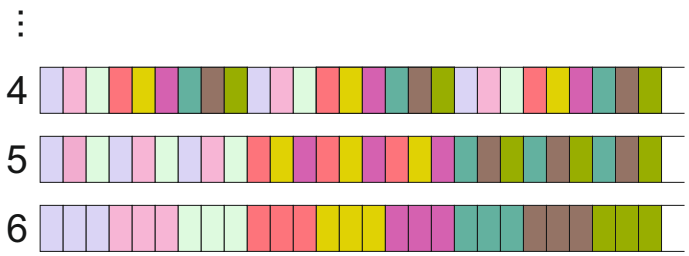
$$\begin{aligned} F(a,b,c) = {} & a^2b^2c + a^2bc^2 - ab^2c^2 + a^2b^2 - a^2bc + a^2c^2 + ab^2c \\ & - a^2c + ab^2 - ac^2 + b^2c + bc^2 - a^2 - b^2 + bc - c^2 - c + 1, \end{aligned}$$

where the calculations are carried out modulo 3 while the residue 2 is represented by the trit -1.

To transform the state W, 27 rounds are performed. A round consists of 6 steps. At each step triplets of trits of W are grouped in a certain way. Then each triplet (a,b,c) is replaced with $S(a,b,c)$.

Groupings are organized as follows (see the picture below). At the first step, the state is divided into 3 words of 243 trits. Trits of these words in the same positions are grouped. In the second step, the state is divided into 9 words of 81 trits. Trits of the 1st, 2nd and 3rd words in the same positions are grouped, then trits of the 4th, 5th and 6th words, and so on. After that, the state is divided into words of length 27, then length 9, then length 3 while maintaining the logic of groupings. In the last sixth step, consecutive triplets of trits are grouped.

Bonus problem (extra scores, a special prize!). Find a collision when the state is initialized in a different way: now W_0, W_2 are not filled with zeros, the word $\underbrace{01\bar{1}01\bar{1}\ldots01\bar{1}}_{243}$ is written in each of them instead.



Groupings (3 last steps, grouped trits are painted the same color)

2.14.2 Solution

For a word u in the alphabet $\mathbf{T}$, let u^m be the word of m copies of u. Supposing $u = u_0u_1\ldots u_{n-1}$ denote $u^{[m]} = u_0^m u_1^m \ldots u_{n-1}^m$. We call a word of the form $u^{[m]}$ *m-fragmented*.

Theorem. Let m be a power of 3, $m \le 729$. The sponge function Curl27-f preserves m-fragmentation, that is, if W is m-fragmented, then Curl27-f(W) is also m-fragmented.

Proof. At the ith step of the Curl27-f round function, the state W is divided into words of length $n = 3^{6-i}$, $i = 1, 2, \ldots, 6$. For $n \le m$ the step function preserves equality of trits inside fragments. It follows from the fact that $S(a,a,a) = (b,b,b)$. For $n > m$ equality is also preserved since in each fragment trits at the different positions are processed in the same way. □

Let m be a small power of 3 (interesting cases are $m = 3, 9, 27$). Consider a ternary string X of length

$$1 + 3 + 3^2 + \ldots + 3^{m-1} = (3^m - 1)/2.$$

The length is given by a word of m ones. Consequently, the initial state of Curl27 when processing X is m-fragmented (one fragment of ones, the remaining fragments of zeros).

Let us choose trits of X so as to preserve m-fragmentation of the state during hashing. This is easy to do using Theorem: each full m-fragment of X must have the form α^m, $\alpha \in \mathbf{T}$, and, in addition, trits of the last (incomplete) fragment must be zero to be consistent with the padding trits. Having achieved m-fragmentation of states, we automatically obtain m-fragmentation of hash values. Now a hash value is determined by $243/m$ trits, each of which is repeated m times. We can find a collision for Curl27 after processing of about $\sqrt{3^{243/m}}$ strings X of the described structure, that is, in time of order

$$3^m \cdot \sqrt{3^{243/m}} = 3^{m+121.5/m}.$$

The minimum of the function above is achieved at $m = 9$. During the attack with $m = 9$ it is required to process approximately $\sqrt{3^{13.5}}$ strings of $9841 = 243 \cdot 40 + 121$ trits each.

An example of colliding messages:

$$X = 0^{243 \cdot 39}(101100110101111100101100000)^{[9]}0^{121},$$
$$X' = 0^{243 \cdot 39}(000011110100111111001000000)^{[9]}0^{121}.$$

This collision was found by Jeremy Jean (National Cybersecurity Agency of France), the only participant who solved the problem.

The preservation of fragmentation is an invariant of Curl27-f which allows to decrease the dimension and thereby effectively solve the basic problem. To solve the bonus problem, Jeremy Jean proposed to use another invariant for Curl27-f: if each part W_0, W_1, W_2 of the state W is 3-expanded, then this fact also holds for Curl27-f(W). Here we call a word $U \in \mathbf{T}^{243}$ 3-*expanded* if it has the form $(abc)^{81}$, $abc \in \mathbf{T}^3$.

In the initial state, the parts W_0 and W_2 are indeed 3-expanded. To comply with the invariant, the part W_1 representing the length of a hashed string X must have one of the forms $(ab1)^{81}$, $(a10)^{81}$ or $(100)^{81}$ (the length is nonzero and positive). As a result, X consists of at least $1+27+\ldots+27^{80} > 3^{240}$ trits.

It is easy to maintain the invariant during hashing: full 243-fragments of X must be 3-expanded and the last incomplete fragment (if it exists) must be filled with zeros. The resulting hash values are 3-expanded, there are only 27 choices for them and a collision will surely be found after processing only 28 strings X. Of course, the attack is impractical: the time of order 3^{240}, which is required only for recording colliding messages, is unacceptably large even compared to the time $3^{243/2}$ of the standard birthday attack.

2.15 Problem "8-bit S-box"

2.15.1 Formulation

Permutations S of the set $\{0,1\}^n$ or $\mathbb{F}_2^n$ are usually called *n-bit S-boxes*. We will focus on the following cryptographic properties of S-boxes:

1. **The (minimal) algebraic degree** of S, denoted by $\deg(S)$, is the minimum of algebraic degrees of all component functions of S.

2. **The nonlinearity of** S, denoted by $\mathrm{nl}(S)$, is the minimal Hamming distance between all component functions of S and the set of all affine functions.
3. **The differential uniformity** of S, denoted by $\mathrm{du}(S)$ is the maximal number of solutions of the equation $S(x) \oplus S(x \oplus \alpha) = \beta$ for any nonzero vector α and any vector β.
4. **The (graph) algebraic immunity** of S, denoted by $\mathrm{ai}(S)$, is the minimal algebraic degree of all nonzero Boolean functions f in $2n$ variables such that $f(x, y) = 0$ for any $x \in \mathbb{F}_2^n$ and $y = S(x)$.

In modern symmetric cryptography, S-boxes of dimension $n = 8$ are probably the most popular. For example, such an S-box is used in the AES block cipher. The characteristics of S_{AES}:

$$(\deg, \mathrm{nl}, \mathrm{du}, \mathrm{ai})(S_{\mathrm{AES}}) = (7, 112, 4, 2).$$

The value $\mathrm{ai}(S_{\mathrm{AES}}) = 2$ means that S_{AES} (and the whole AES) can be compactly described by quadratic equations. This can be a weakness in the context of algebraic attacks.

Imposing the restrictions $(\deg, \mathrm{ai})(S) = (7, 3)$ (optimal values), we need to maximize $\mathrm{nl}(S)$ and minimize $\mathrm{du}(S)$. The current best result [7, 8] is

$$(\deg, \mathrm{nl}, \mathrm{du}, \mathrm{ai})(S) = (7, 108, 6, 3).$$

Problem for a special prize! You need to improve this result: find 8-bit S with $\mathrm{nl}(S) > 108$ and/or $\mathrm{du}(S) < 6$ while preserving $\deg(S) = 7$ and $\mathrm{ai}(S) = 3$.

Remarks. Let us recall relevant definitions.

1. A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ can be uniquely represented in the *algebraic normal form* (ANF) in the following way: $f(x) = \bigoplus_{I \in \mathcal{P}(N)} a_I \big(\prod_{i \in I} x_i \big)$, where $\mathcal{P}(N)$ is the power set of $N = \{1, \ldots, n\}$ and $a_I \in \mathbb{F}_2$.
2. The *algebraic degree* of F is degree of its ANF: $\deg(F) = \max\{|I| : \ a_I \neq 0, \ I \in \mathcal{P}(N)\}$.
3. Boolean functions of the algebraic degree not more than 1 are called *affine*.
4. The Hamming distance between Boolean functions f and g is the number of vectors $x \in \mathbb{F}_2^n$ such that $f(x) \neq g(x)$.
5. A function $S : \mathbb{F}_2^n \to \mathbb{F}_2^n$ can be given as $S = (s_1, \ldots, s_n)$, where s_i is a Boolean function; a nontrivial linear combination of $s_1, \ldots, s_n$ is a *component* function of S.

2.15.2 Solution

There were no valuable ideas from the Olympiad participants. The problem remains unsolved for the considered configuration of cryptographic properties. There exist several dozen of constructions, based on well-known butterfly structure, that provide current record $(7, 108, 6, 3)$, see [7, 8]. This leads to the idea that if candidates for improvement exist, then they are likely outside the known structures and constructions of cryptographic permutations.

2.16 Problem "Conjecture"

2.16.1 Formulation

Let $\mathbb{F}_2$ be the finite field with two elements and n be any positive integer larger than or equal to 3. Let $f(X)$ be an irreducible polynomial of degree n over $\mathbb{F}_2$. It is known that the set of the equivalence classes β of polynomials over $\mathbb{F}_2$ modulo $f(X)$ is a finite field of order 2^n, that we shall denote by $\mathbb{F}_{2^n}$. It is known that different choices of the irreducible polynomial give automorphic

finite fields and such choice has then no incidence on the algebraic problems on the corresponding fields.

Problem for a special prize! Prove or disprove the following

Conjecture. Let k be co-prime with n. For every $\beta \in \mathbb{F}_{2^n}$, let $F(\beta) = \beta^{4^k-2^k+1}$. Let $\Delta = \{F(\beta) + F(\beta+1) + 1;\ \beta \in \mathbb{F}_{2^n}\}$. For every distinct nonzero v_1, v_2 in $\mathbb{F}_{2^n}$, we have

$$|\{(x, y, z) \in \Delta^3;\ v_1x + v_2y + (v_1 + v_2)z = 0\}| = 2^{2n-3}.$$

Example for $n = 3$: we can take $f(X) = X^3 + X + 1$, then each element β of the field $\mathbb{F}_{2^3}$ can be written as a polynomial of degree at most 2: $a_0 + a_1X + a_2X^2$, with $a_0, a_1, a_2 \in \mathbb{F}_2$. The element 0 corresponds to the null polynomial; and the unity, denoted by 1, corresponds to the constant polynomial 1. We can calculate the table of multiplication in $\mathbb{F}_{2^3}$ (the table of addition just corresponds to adding polynomials of degree at most 2); this allows calculating any power of any element of the field and check the property.

2.16.2 Solution

This mathematical problem is open and difficult. It was presented in [5] for the first time and discussed in [6]. The conjecture was verified for small n (odd values $n \leqslant 11$, even values $n \leqslant 8$). The Olympiad participants suggested several ideas. Unfortunately, none of them gave significant advances to prove a conjecture or search for a counterexample. The team of Kristina Geut, Sergey Titov, and Dmitry Ananichev (Ural State University of Railway Transport) and the team of Alexey Zelenetskiy, Mikhail Kudinov, and Denis Nabokov (Bauman Moscow State Technical University) proved the conjecture for a particular case $k = 1$. Nevertheless, this case is peculiar since the function is then quadratic and the result is known for quadratic functions. The proofs cannot be generalized to the common case.

3 Winners of the Olympiad

Here we list information about the winners of NSUCRYPTO'2019 in Tables 9, 10, 11, 12, 13.

Table 9: **Winners of the first round in school section A ("School Student")**

Place	Name	Country, City	School	Scores
1	Borislav Kirilov	Bulgaria, Sofia	The First Private Mathematical Gymnasium	16
1	Alexey Lvov	Russia, Novosibirsk	Gymnasium 6	16
2	Lenart Bucar	Slovenia, Ljubljana	Gymnasium Bezigrad	15
3	Varvara Lebedinskaya	Russia, Novosibirsk	The Specialized Educational Scientific Center of Novosibirsk State University	14
3	Gabriel Ericson	Sweden, Örebro	Tullangsskolan	14
Diploma	Vlad Coneschi	Romania, Slatina	Radu Greceanu National College	11
Diploma	Wang Duanyu	Singapore, Singapore	New Town Primary School	9
Diploma	Vlad Ratnikov	Russia, Yaroslavl	School 33 of Yaroslavl	9
Diploma	Nikita Kukin	Russia, Moscow	Gymnasium 1540 of Moscow	8
Diploma	Michail Kostochka	Russia, Novosibirsk	Lyceum 130	8

Table 10: **Winners of the first round, section B (in the category "University Student")**

Place	Name	Country, City	University	Scores
1	Maxim Plushkin	Russia, Moscow	Lomonosov Moscow State University	22
1	Mikhail Kudinov	Russia, Moscow	Bauman Moscow State Technical University	21
2	Narendra Patel	India, Roorkee	Indian Institute of Technology Roorkee	19
2	Vladimir Schavelev	Russia, Saint Petersburg	Saint Petersburg State University	19
3	Thanh Nguyen Van	Vietnam, Ho Chi Minh City	Ho Chi Minh City University of Technology	16
3	Daria Grebenchuk	Russia, Yaroslavl	Yaroslavl State University	16
3	Roman Gibadulin	Russia, Yaroslavl	Yaroslavl State University	16
3	Tuong Nguyen	Vietnam, Ho Chi Minh City	Ho Chi Minh City University of Technology	15
Diploma	Denis Nabokov	Russia, Moscow	Bauman Moscow State Technical University	14
Diploma	Filip Dashtevski	Macedonia, Kumanovo	TU Delft	14
Diploma	Sayooj Samuel	India, Kollam	Amrita University	14
Diploma	Paul Cotan	Romania, Iaşi	Alexandru Ioan Cuza University	13
Diploma	Karina Kruglik	Belarus, Minsk	Belarusian State University	13
Diploma	Hosein Hadipour	Iran, Tehran	University of Tehran	13
Diploma	Polina Raspopova	Russia, Yekaterinburg	Ural State University of Railway Transport	12
Diploma	Gorazd Dimitrov	Macedonia, Skopje	Ecole Polytechnique	12
Diploma	Diana Bespechnaya	Russia, Moscow	Bauman Moscow State Technical University	12
Diploma	Nikolay Prudkovskiy	Russia, Moscow	Bauman Moscow State Technical University	12
Diploma	Riccardo Zanotto	Italy, Pisa	University of Pisa	12
Diploma	Dmitry Zakharov	Russia, Moscow	National Research Nuclear University MEPhI	12

Table 11: **Winners of the first round, section B (in the category "Professional")**

Place	Name	Country, City	Organization	Scores
1	Henning Seidler	Germany, Berlin	TU Berlin	26
2	Samuel Tang	Hong Kong, Hong Kong	Black Bauhinia	20
2	Madalina Bolboceanu	Romania, Bucharest	Bitdefender	20
3	Irina Slonkina	Russia, Moscow	National Research Nuclear University MEPhI	16
Diploma	Harry Lee	Hong Kong, Hong Kong	Blocksquare Limited	14
Diploma	Alexey Chilikov	Russia, Moscow	Bauman Moscow State Technical University	14
Diploma	Victoria Vlasova	Russia, Moscow	Bauman Moscow State Technical University	14
Diploma	Darko Ninkovic	Serbia, Belgrade	University of Belgrade	13
Diploma	Dheeraj M Pai	India, Chennai	Hyperweb Media Private Limited	13
Diploma	Dmitry Ananichev	Russia, Yekaterinburg	Ural Federal University	13
Diploma	Ekaterina Kulikova	Germany, Munich		13
Diploma	George Teseleanu	Romania, Bucharest	Institute of Mathematics of the Romanian Academy	12

Table 12: **Winners of the second round (in the category "University student")**

Place	Name	Country, City	University	Scores
1	Alexey Zelenetskiy, Mikhail Kudinov, Denis Nabokov	Russia, Moscow	Bauman Moscow State Technical University	51
2	Ngoc Ky Nguyen, Dung Truong, Phuoc Nguyen Ho Minh	Vietnam, Ho Chi Minh City; France, Paris	Ho Chi Minh City University of Technology, Ecole Normale Superieure	43
2	Thanh Nguyen Van, Quoc Bao Nguyen, Ngan Nguyen	Vietnam, Ho Chi Minh City	Ho Chi Minh City University of Technology	40
3	Maxim Plushkin	Russia, Moscow	Lomonosov Moscow State University	34
3	Ilya Trusevich, Maxim Bibik, Alexander Shulga	Belarus, Minsk	Belarusian State University	38
Diploma	Paul Cotan, Evgnosia-Alexandra Kelesidis	Romania, Iaşi	Alexandru Ioan Cuza University	26
Diploma	Roman Sychev, Diana Bespechnaya, Nikolay Prudkovskiy	Russia, Moscow	Bauman Moscow State Technical University	24
Diploma	Vladimir Paprotski, Dmitry Zarembo, Karina Kruglik	Belarus, Minsk	Belarusian State University	21
Diploma	Vitaliy Cherkashin, Zoya Tabikhanova, Evgenia Bykova	Russia, Novosibirsk	Novosibirsk State Pedagogical University	18

Table 13: **Winners of the second round (in the category "Professional")**

Place	Names	Country, City	Organization	Scores
1	Irina Slonkina, Mikhail Sorokin, Vladimir Bobrov	Russia, Moscow	Bauman Moscow State Technical University	48
1	Kristina Geut, Sergey Titov, Dmitry Ananichev	Russia, Yekaterinburg	Ural State University of Railway Transport, Ural Federal University	46
2	Henning Seidler, Katja Stumpp	Germany, Berlin	Berlin Technical University	42
3	Victoria Vlasova, Mikhail Polyakov, Alexey Chilikov	Russia, Moscow	Bauman Moscow State Technical University	37
3	Duc Tri Nguyen, Quan Doan, Tuong Nguyen	Vietnam, Ho Chi Minh City	Cryptographic Engineering Research Group, pwnphofun, Ho Chi Minh City University of Technology	36
3	Madalina Bolboceanu, Andrei Mogage, Radu Titiu	Romania, Bucharest	Bitdefender, Alexandru Ioan Cuza University	34
Diploma	Elena Kirshanova, Semyon Novoselov, Nikita Kolesnikov	Russia, Kaliningrad	Immanuel Kant Baltic Federal University	28
Diploma	Vyacheslav Salmanov, Evgeniya Ishchukova, Nikita Kutovoy	Russia, Taganrog	Southern Federal University	22
Diploma	Jeremy Jean	France, Paris	National Cybersecurity Agency of France	20
Diploma	Khai Hanh Tang, Pham Phuong, Yi Tu	Singapore, Singapore	Nanyang Technological University	21
Diploma	Harry Lee, Samuel Tang	Hong Kong, Hong Kong	Black Bauhinia	20
Diploma	Danh Nam Tran, Thu Hien Chu Thi, Phu Nghia Nguyen	Vietnam, Ho Chi Minh City	Ho Chi Minh City Pedagogical University, Japan Advanced Institute of Science and Technology, Ho Chi Minh City University of Technology	20

References

[1] Agievich S., Gorodilova A., Idrisova V., Kolomeec N., Shushuev G., Tokareva N. Mathematical problems of the second international student's Olympiad in cryptography. Cryptologia. 2017, V. 41, No. 6, pp. 534–565.

[2] Agievich S., Gorodilova A., Kolomeec N., Nikova S., Preneel B., Rijmen V., Shushuev G., Tokareva N., Vitkup V. Problems, solutions and experience of the first international student's Olympiad in cryptography. Prikladnaya Diskretnaya Matematika (Applied Discrete Mathematics). 2015, No. 3, pp. 41–62.

[3] Brinkmann M., Leander G. On the classification of APN functions up to dimension five. Designs, codes and cryptography. 2008, V. 49, pp. 273–288.

[4] De Canni‘ere C. "Analysis and Design of Symmetric Encrytption Algorithms," Ph.D. thesis, 2007.

[5] Carlet C. Componentwise APNness, Walsh uniformity of APN functions, and cyclic-additive difference sets. Finite Fields and Their Applications. 2018, V. 53, pp. 226–253.

[6] Carlet C. On APN exponents, characterizations of differentially uniform functions by the Walsh transform, and related cyclic-difference-set-like structures. Proceedings of WCC 2017. Designs, Codes and Cryptography (Postproceedings of WCC 2017). V. 87 (2), pp. 203–224, 2018.

[7] de la Cruz Jimènez R. A. Generation of 8-Bit S-Boxes Having Almost Optimal Cryptographic Properties Using Smaller 4-Bit S-Boxes and Finite Field Multiplication. In: Lange T., Dunkelman O. (eds) Progress in Cryptology – LATINCRYPT 2017. LNCS, 2019, V. 11368, pp. 191–206.

[8] Fomin D. B. New classes of 8-bit permutations based on a butterfly structure. Math. vopr. kript. 2019, V. 10(2), pp. 169–180. `https://ctcrypt.ru/files/files/2018/09_Fomin.pdf`.

[9] Geut K., Kirienko K., Sadkov P., Taskin R., Titov S. On explicit constructions for solving the problem "A secret sharing". Prikladnaya Diskretnaya Matematika. Prilozhenie. 2017, No. 10, pp. 68–70 (in Russian).

[10] Gorodilova A., Agievich S., Carlet C., Gorkunov E., Idrisova V., Kolomeec N., Kutsenko A., Nikova S., Oblaukhov A., Picek S., Preneel B., Rijmen V., Tokareva N. Problems and solutions of the Fourth International Students Olympiad in Cryptography (NSUCRYPTO). Cryptologia. 2019, V. 43, I. 2, pp. 138–174.

[11] Gorodilova A., Agievich S., Carlet C., Hou X., Idrisova V., Kolomeec N., Kutsenko A., Mariot L., Oblaukhov A., Picek S., Preneel B., Rosie R., Tokareva N. The Fifth International Students' Olympiad in Cryptography - NSUCRYPTO: problems and their solutions. Cryptologia. 2020, V. 44, I. 3, pp. 223–256.

[12] Lewand R. E. *Cryptological Mathematics,* MAA, Washington, 2000.

[13] Schneier B. Applied Cryptography: Protocols, Algorithms and Source Code in C. Wiley; 2nd edition, 1996.

[14] Tokareva N., Gorodilova A., Agievich S., Idrisova V., Kolomeec N., Kutsenko A., Oblaukhov A., Shushuev G. Mathematical methods in solutions of the problems from the Third International Students' Olympiad in Cryptography. Prikladnaya Diskretnaya Matematika (Applied Discrete Mathematics). 2018, No. 40, pp. 34–58.

[15] https://nsucrypto.nsu.ru/

[16] https://nsucrypto.nsu.ru/unsolved-problems/

[17] https://nsucrypto.nsu.ru/archive/2019/round/2/task/4/

[18] https://nsucrypto.nsu.ru/media/Olympiads/2019/Round_2/Tasks/curl27.java

[19] Find Words Using Pattern Matching, in *Litscape.com*. Available at http://www.litscape.com/word_tools/pattern_match.php.

[20] Letter Frequency, in *Wikipedia*. Available at https://en.wikipedia.org/wiki/Letter_frequency.

[21] https://www.ibm.com/blogs/research/2018/01/quantum-prizes/

Metrical properties of the set of bent functions in view of duality *

Aleksandr Kutsenko[1,2], Natalia Tokareva[1]

[1]Sobolev Institute of Mathematics, Novosibirsk, Russia
[2]Novosibirsk State University, Novosibirsk, Russia

Email: alexandrkutsenko@bk.ru, tokareva@math.nsc.ru

Abstract

In this work we give a review of metrical properties of the entire set of bent functions and its significant subclasses of self-dual and anti-self-dual bent functions. We give results for iterative contruction of bent functions in $n+2$ variables based on the concatenation of four bent functions and consider related open problem proposed by one of the authors. Criterion of self-duality of such functions is discussed. It is explored that the pair of sets of bent functions and affine functions as well as a pair of sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables is a pair of mutually maximally distant sets that implies metrical duality. Groups of automorphisms of the sets of bent functions and (anti-)self-dual bent functions are discussed. The solution to the problem of preserving bentness and the Hamming distance between bent function and its dual within automorphisms of the set of all Boolean functions in n variables is considered.

Keywords — Boolean bent function, self-dual bent function, Hamming distance, metrical regularity, automorphism group, iterative construction

1 Introduction

How much do we know about some cryptographic objects? One way to measure it is to describe what we can do with them. Otherwise to characterize groups of automorphisms of these objects — separately for each object or together while they form some special class. The question about the group of automorphisms of a set in the Boolean cube necessarily leads us to metrical properties of this set.

That is why we are very interested in *metrical properties* of distinct cryptographic Boolean functions.

The term "bent function" was introduced by Oscar Rothaus in the 1960s [31]. It is known [39], that at the same time Boolean functions with maximal nonlinearity were also studied in the Soviet Union. The term *minimal function*, which is actually a counterpart of a bent function, was proposed by the Soviet scientists Eliseev and Stepchenkov in 1962.

Bent functions have connections with such combinatorial objects as Hadamard matrices and difference sets. Since bent functions have maximum Hamming distance to linear structures and affine functions they deserve attention for practical applications in symmetric cryptography, in particular, for block and stream ciphers. We refer to the survey [5] and monographies of S. Mesnager [26] and N. Tokareva [39] for more information concerning known results and open problems related to bent functions. Results regarding the study of metrical properties, in particular, distances between bent functions, one can find in article [17].

*The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

In this paper we give review on metrical properties of the entire class of bent function $\mathcal{B}_n$ and its important subclasses — self-dual bent functions $\mathrm{SB}^+(n)$ (i.e. functions such that $f = \widetilde{f}$) and anti-self-dual bent functions $\mathrm{SB}^-(n)$ (i.e. functions such that $f \oplus 1 = \widetilde{f}$), where $\widetilde{f}$ is the dual of f. We suppose that the *keys* to the nontrivial and important properties of the class of bent functions are in understanding how does the *duality mapping* $f \to \widetilde{f}$ operate with bent functions. Recall that $\widetilde{\widetilde{f}} = f$ for every bent function f. It is important to note that the duality mapping is the *unique* known isometric mapping of the bent functions into themselves that can not be extended to a typical isometry of the whole set of all Boolean functions that preserves bent functions.

On other hand, the essence of bent functions is expressed in their metrical properties, namely in maximizing distances between them and affine functions. Note that this very idea in more general form is realized in the concept of metrical complement and metrically regular sets. Recall that $\widehat{X}$ is the metrical complement of the set of functions X if it contains all Boolean functions that are on the maximal possible distance from X. The set is metrically regular, if $\widehat{\widehat{X}} = X$. There is a some similarity to the self-duality of bent functions, is not it?

Our attention is drawn to automorphism groups of the sets $\mathcal{B}_n$, $\mathcal{A}_n$, $\mathrm{SB}^+(n)$, $\mathrm{SB}^-(n)$ and their metrical properties. Previously, we established that the set of all bent functions $\mathcal{B}_n$ and the set of all affine functions $\mathcal{A}_n$ form a pair of metrically regular sets, i.e. $\widehat{\widehat{\mathcal{B}_n}} = \widehat{\mathcal{A}_n} = \mathcal{B}_n$. Now we prove the same fact for the classes of self-dual and anti-self-dual functions: they form another such pair of metrically complement functions, i.e. $\widehat{\widehat{\mathrm{SB}^+(n)}} = \widehat{\mathrm{SB}^-(n)} = \mathrm{SB}^+(n)$. In both cases for elements in a pair of metrically regular sets we prove the coincidence of automorphism groups. Thus, $\mathrm{Aut}\,(\mathcal{B}_n) = \mathrm{Aut}\,(\mathcal{A}_n)$ and $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right)$. Some other curious properties of bent functions related to their special constructions are discussed in the paper.

The paper has the following structure: notation and definitions are in the Section 2. In Section 3 the duality of a bent function is described, including some its important properties and relevant hypothesis proposed by one of the authors (Section 3.1). Some general and metrical properties of the set of bent functions which coincide with their duals, namely self-dual bent functions, are given in Section 3.2. In Section 4 we discuss the iterative construction of bent function in $n + 2$ variables based on the concatenation of four bent functions in n variables. The lower bounds on its cardinality and open problem relevant for the set of bent function are in Section 4.1. Criterion of self-duality for bent iterative functions and its corollaries for sign functions together with constructions of self-dual bent functions are discussed in Sections 4.2 and 4.3. In Section 5 the metrical complement of the set of bent functions is studied (Section 5.2) and the results regarding metrical regularity of the set of bent functions and the set of affine functions are given. Metrical complement of the set of (anti-)self-dual bent functions is in Section 5.3. In Section 6 groups of automorphisms of considered sets are studied. The group of automorphisms of the set of bent functions is characterized in Section 6.3 while the (anti-)self-dual case is in Section 6.4. In Section 7 we consider some relations between isometric mappings and the duality of bent function. Isometric mappings which define bijections between the sets of self-dual and anti-self dual bent functions are described in Section 7.1. The Rayleigh quotient of a Boolean function and description of isometric mappings that perserve it or change it for every Boolean function is in Section 7.2. The meaning of the Rayleigh quotient in a scope of bent functions is discussed as well.

2 Notation

Let $\mathbb{F}_2^n$ be a space of binary vectors of length n. Denote, following [13], the orthogonal group of index n over the field $\mathbb{F}_2$ as

$$\mathcal{O}_n = \left\{ L \in \mathrm{GL}\,(n, \mathbb{F}_2) : LL^{\mathrm{T}} = I_n \right\},$$

where L^{T} denotes the transpose of L and I_n is an identical matrix of order n over the field $\mathbb{F}_2$.

A *Boolean function* f in n variables is a map from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. Its *sign function* is $F(x) = (-1)^{f(x)}$, $x \in \mathbb{F}_2^n$. We will also refer to a sign function as to a vector from the set $\{\pm 1\}^{2^n}$:

$$F = (-1)^f = \left((-1)^{f_0}, (-1)^{f_1}, \ldots, (-1)^{f_{2^n-1}}\right) \in \{\pm 1\}^{2^n},$$

where $(f_0, f_1, \ldots, f_{2^n-1}) \in \mathbb{F}_2^{2^n}$ is a truth-table representation of f with arguments given in the lexicographic order. The set of Boolean functions in n variables is denoted by $\mathcal{F}_n$.

The *algebraic normal form* (ANF, Zhegalkin polynomial) of a Boolean function $f \in \mathcal{F}_n$ is defined to be

$$f(x_1, x_2, \ldots, x_n) = \bigoplus_{(i_1, i_2, \ldots, i_n) \in \mathbb{F}_2^n} a_{i_1 i_2 \ldots i_n} x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n},$$

where $a_z \in \mathbb{F}_2$ for any $z \in \mathbb{F}_2^n$ (with the convention $0^0 = 1$). The *algebraic degree* $\deg(f)$ of a Boolean function f is the maximal degree of monomials which occur in its algebraic normal form with nonzero coefficients.

The *Hamming weight* $\mathrm{wt}(x)$ of the vector $x \in \mathbb{F}_2^n$ is the number of nonzero coordinates of x. The *Hamming weight* $\mathrm{wt}(f)$ of the function $f \in \mathcal{F}_n$ is the Hamming weight of its vector of values. The *Hamming distance* $\mathrm{dist}(f, g)$ between Boolean functions f, g in n variables is a cardinality of the set $\{x \in \mathbb{F}_2^n : f(x) \oplus g(x) = 1\}$. For $x, y \in \mathbb{F}_2^n$ denote $\langle x, y \rangle = \bigoplus_{i=1}^{n} x_i y_i$. Boolean functions in n variables of the form $f(x) = \langle a, x \rangle \oplus a_0$, $x \in \mathbb{F}_2^n$, where $a_0 \in \mathbb{F}_2$, $a \in \mathbb{F}_2^n$, are called *affine* functions. The set of all affine functions in n variables is denoted by $\mathcal{A}_n$.

The *Walsh "— Hadamard transform* (WHT) of a Boolean function f in n variables is an integer valued function $W_f : \mathbb{F}_2^n \to \mathbb{Z}$, defined as

$$W_f(y) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus \langle x, y \rangle}, \quad y \in \mathbb{F}_2^n.$$

A Boolean function f in an even number n of variables is called *bent* if

$$|W_f(y)| = 2^{n/2}$$

for all $y \in \mathbb{F}_2^n$. The set of all bent functions in n variables is denoted by $\mathcal{B}_n$.

3 The dual of a bent function

From the definition of a bent function it follows that for any $y \in \mathbb{F}_2^n$ we have

$$W_f(y) = (-1)^{\widetilde{f}(y)} 2^{n/2}$$

for some $\widetilde{f} \in \mathcal{F}_n$. The Boolean function $\widetilde{f}$ defined above is called the *dual* function of the bent function f. Thus, for any bent function in n variables its dual Boolean function is uniquely defined. The duality of bent functions was introduced by Dillon [11].

3.1 Properties

Some basic known properties of dual functions are the following [3]:

- Every dual function is a bent function;
- If $\widetilde{f}$ is dual to f and $\widetilde{\widetilde{f}}$ is dual to $\widetilde{f}$, then $\widetilde{\widetilde{f}} = f$;
- The mapping $f \to \widetilde{f}$ which acts on the set of bent functions, preserves the Hamming distance.

There is the following connection between the algebraic degrees of a bent function and its dual [14]:

$$n/2 - \deg(f) \geqslant \frac{n/2 - \deg\left(\widetilde{f}\right)}{\deg\left(\widetilde{f}\right) - 1}.$$

Some results obtained for dual functions can be used in proving the results concerning bent functions, in particular, the connection between ANF coefficients of a bent function and its dual, see [8]:

$$\sum_{x \preccurlyeq y} f(x) = 2^{\mathrm{wt}(y)} - 2^{n/2-1} + 2^{\mathrm{wt}(y)-n/2} \sum_{x \preccurlyeq y \oplus 1} \widetilde{f}(x).$$

One of the most important problem in bent functions is to find the number of them. A new approach to this problem was introduced in [35], see Section 4.1, and the following hypothesis was formulated.

Hypothesis (Tokareva, 2011). Any Boolean function in n variables of degree not more than $n/2$ can be represented as the sum of two bent functions in n variables, where $n \geqslant 2$ is an even number.

The review of partial results regarding this problem and also in favour of the Hypothesis one can find in [37]. It was also proved in [38] that

Theorem 1 ([38]). *A bent function in $n \geqslant 4$ variables can be represented as the sum of two bent functions in n variables if and only if its dual bent function does.*

So, it follows that the mentioned Hypothesis with the decomposition problem, see Section 4.1, can not be considered separately for a bent function and its dual.

It is worth noting that this Hypothesis is a counterpart of the Goldbach's conjecture in number theory unsolved since 1742: any even number $n > 4$ can be represented as the sum of two prime numbers.

Isometric mappings of the set of all Boolean functions in n variables to itself which preserve bentness and the Hamming distance between every bent function and its dual were characterized in [20], namely it was proved that

Theorem 2 ([20]). *An isometric mapping φ of the set of all Boolean functions in n variables into itself preserves bentness and the Hamming distance between every bent function and its dual if and only if φ has form*

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x \rangle \oplus d, \quad x \in \mathbb{F}_2^n, \tag{1}$$

for some $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$.

3.2 Self-duality

If a bent function f coincides with its dual it is said to be *self-dual*, that is $f = \widetilde{f}$. A bent function which coincides with the negation of its dual is called an *anti-self-dual*, that is $f = \widetilde{f} \oplus 1$. The set of (anti-)self-dual bent functions in n variables, according to [15], is denoted by $\mathrm{SB}^+(n)$ $\left(\mathrm{SB}^-(n)\right)$.

Self-dual bent functions were explored in paper of C. Carlet et al. [4] in 2010, where some properties and constructions were given. All equivalence classes of self-dual bent functions in 2, 4, and 6 variables and all quadratic self-dual bent functions in 8 variables with respect to a restricted form of an affine transformation (1), which preserves self"=duality, were also presented. Further, equivalence classes of cubic self-dual bent functions in 8 variables with respect to the mentioned above restricted form of affine transformation one can find in [12]. In [15], a classification of quadratic self-dual bent functions was obtained. The upper bound

for the cardinality of the set of self-dual bent functions was given in [16]. In [21, 25] one can find new constructions of self-dual bent functions. A connection of quaternary self-dual bent functions and self-dual bent Boolean functions was shown in [32]. In [19] it was proved that for any $d \in \{2, 3, \ldots, n/2\}$ there exists a self-dual bent function of algebraic degree d.

In papers [18, 19, 20] metrical properties of the sets of (anti-)self-dual bent functions in n variables were studied. Below we briefly discuss some of them.

Recall that bent functions in $2k$ variables which have a representation

$$f(x, y) = \langle x, \pi(y) \rangle \oplus g(y), \quad x, y \in \mathbb{F}_2^k,$$

where $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and g is a Boolean function in k variables, form the well known *Maiorana–McFarland class* of bent functions [24]. Let the denotion $\mathrm{SB}^+_{\mathcal{M}}(n)$ stands for the set of self-dual Maiorana–McFarland bent functions and $\mathrm{SB}^-_{\mathcal{M}}(n)$ for the set of anti-self-dual ones both in n variables. Necessary and sufficient conditions of (anti-)self-duality of bent functions from Maiorana–McFarland class are known from [4], namely a Maiorana–McFarland bent function $f(x, y) \in \mathcal{F}_{2k}$ is self-dual bent if and only if

$$\pi(y) = L(y \oplus c), \quad g(y) = \langle c, y \rangle \oplus d, \quad y \in \mathbb{F}_2^k,$$

where $L \in \mathcal{O}_k$, $c \in \mathbb{F}_2^k$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2^n$. Note that $\left|\mathrm{SB}^+_{\mathcal{M}}(2k)\right| = 2^k \cdot |\mathcal{O}_k|$. In [18] the set of possible Hamming distance between such self-dual bent functions was found,

Theorem 3 ([18]). *Let $n \geqslant 4$ and $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n) \cup \mathrm{SB}^-_{\mathcal{M}}(n)$, then*

$$\mathrm{dist}(f, g) \in \left\{ 2^{n-1}, 2^{n-1}\left(1 \pm \frac{1}{2^r}\right), r = 0, 1, ..., n/2 - 1 \right\}.$$

Moreover, if either $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n)$ or $f, g \in \mathrm{SB}^-_{\mathcal{M}}(n)$, then all distances except 2^{n-1} are attainable, and for any pair $f \in \mathrm{SB}^+_{\mathcal{M}}(n)$ and $g \in \mathrm{SB}^-_{\mathcal{M}}(n)$ it holds $\mathrm{dist}(f, g) = 2^{n-1}$.

By analysis of the set of distances from Theorem 3 the minimal Hamming distance between considered functions can be obtained:

Corollary 1. *The minimal Hamming distance between (anti-)self-dual Maiorana–McFarland bent functions in $n \geqslant 4$ variables is equal to 2^{n-2}.*

Moreover, since the minimal Hamming distance between quadratic Boolean functions in n variables (which correspond to codewords of the $\mathrm{RM}(2, n)$ code) is at least 2^{n-2} [22], the following fact holds

Corollary 2. *If $n \geqslant 4$, then the minimal Hamming distance between quadratic bent functions can be attained on (anti-)self-dual Maiorana–McFarland bent functions.*

It is known that the minimal Hamming distance between bent functions in n variables is $2^{n/2}$ [17]. In [19] it was proved that this extremal value can be attained on (anti-)self-dual bent functions.

Theorem 4 ([19]). *Let $n \geqslant 4$, then the minimal Hamming distance between distinct (anti-)self-dual bent functions in n variables is equal to $2^{n/2}$.*

4 Iterative construction $\mathcal{BI}$

Let f_0, f_1, f_2, f_3 be Boolean functions in n variables. Consider a Boolean function g in $n + 2$ variables which is defined as

$$g(00, x) = f_0(x), \quad g(01, x) = f_1(x), \quad g(10, x) = f_2(x), \quad g(11, x) = f_3(x), \quad x \in \mathbb{F}_2^n.$$

It is known (Preneel et al., 1991; see also [1, 35]) that under condition $f_0, f_1, f_2, f_3 \in \mathcal{B}_n$ the mentioned function g is a bent function in $n+2$ variables if and only if

$$\widetilde{f_0} \oplus \widetilde{f_1} \oplus \widetilde{f_2} \oplus \widetilde{f_3} = 1,$$

that gives the construction of a bent function in $n+2$ variables through the concatenation of vectors of values of four bent functions in n variables [29].

Following N. Tokareva [35], we will refer to bent functions obtained by this construction as *bent iterative functions* ($\mathcal{BI}$) and denote the set of such bent functions in n variables by $\mathcal{BI}_n$.

In [6], the comparison of cardinalities of different known iterative constructions of bent functions in $n \leqslant 10$ variables was presented and the class $\mathcal{BI}$ had the biggest cardinality among them.

According to [1], there exist bent functions from Maiorana–McFarland class [24] and from the class $\mathcal{PS}$ (Partial Spreads) [11] that can not be represented as bent iterative functions. Also, from paper [2] on nonnormal bent functions, it follows that there exist bent functions in $\mathcal{BI}_n$ that are nonequivalent to Maiorana–McFarland bent functions.

4.1 Lower bounds on the cardinality and related open problem

In paper [35] some possible ways of how to calculate the number of bent iterative functions were shown.

Theorem 5 ([35]). *For any even $n \geqslant 4$*

$$|\mathcal{BI}_n| = \sum_{f' \in \mathcal{B}_{n-2}} \sum_{f'' \in \mathcal{B}_{n-2}} |(\mathcal{B}_{n-2} \oplus f') \cap (\mathcal{B}_{n-2} \oplus f'')| \,.$$

Denote $X_n = \{f \oplus h : f, h \in \mathcal{B}_n\}$ and consider the system $\{C_f : f \in \mathcal{B}_n\}$ of its subsets defined as $C_f = \mathcal{B}_n \oplus f$. So,

$$X_n = \bigcup_{f \in \mathcal{B}_n} C_f.$$

Let ψ be an element of X_n. The number of subsets C_f that cover ψ, according to [35], is called *multiplicity* of ψ and is denoted by $m(\psi)$. One can notice that if ψ is covered by C_f, then it is covered by any set $C_{f'}$, where f' is obtained from f by adding an affine function.

In [35], the exact number of bent iterative functions through the multiplicities was obtained.

Theorem 6 ([35]). *For any even $n \geqslant 2$*

$$|\mathcal{BI}_{n+2}| = \sum_{\psi \in C_f} m^2(\psi) \,.$$

So, in order to evaluate $|\mathcal{BI}_{n+2}|$ (and then $|\mathcal{B}_{n+2}|$) we have to study the set X_n and the distribution of multiplicities for its elements. Such an analysis, as shown in [35], gives the following lower bound.

Theorem 7 ([35]). *For any even $n \geqslant 2$*

$$\frac{|\mathcal{B}_{n+2}|^4}{|X_n|} \leqslant |\mathcal{BI}_{n+2}| \leqslant |\mathcal{B}_{n+2}| \,.$$

Thus, for calculating the exact number of bent iterative functions, one has to study the structure of the set X_n. So, we come to a new problem statement.

Open problem: bent sum decomposition (Tokareva, 2011). What Boolean functions can be represented as the sum of two bent functions in n variables? How many such representations does a Boolean function admit?

The related Hypothesis was previuosly mentioned in the Section 3.1.

4.2 Self-dual bent iterative functions

The set of (anti-)self-dual bent functions from $\mathcal{BI}_n$ is further denoted by $\mathrm{SB}^+_{\mathcal{BI}}(n)$ $\left(\mathrm{SB}^-_{\mathcal{BI}}(n)\right)$.

In paper [19], the necessary and sufficient conditions of self-duality of bent iterative functions were studied, namely the following result was obtained. Namely, by taking constant function h one can obtain two constructions of self-dual bent iterative functions in $n+2$ variables:

Theorem 8 ([19]). *Let $g \in \mathcal{BI}_{n+2}$. Then g is self-dual bent if and only if there exists such pair of functions $g_1, g_2 \in \mathcal{B}_n$:*

$$\begin{aligned} f_0 &= (g_1 \oplus g_2)\, h \oplus g_1 = \widetilde{g}_2, \\ f_1 &= (g_1 \oplus g_2)\, h \oplus g_2 = \widetilde{g_1 \oplus h}, \\ f_2 &= (g_1 \oplus g_2)\, h \oplus g_2 \oplus h = \widetilde{g}_1, \\ f_3 &= (g_1 \oplus g_2)\, h \oplus g_1 \oplus h \oplus 1 = \widetilde{g_2 \oplus h} \oplus 1, \end{aligned}$$

where the function $h \in \mathcal{F}_n$ is uniquely defined by a pair of bent functions g_1, g_2, namely:

$$h = g_1 \oplus \widetilde{g}_1 \oplus g_2 \oplus \widetilde{g}_2.$$

Two iterative constructions of self-dual bent functions immeditely follow from Theorem 8, as it was shown in [19].

Corollary 3. *Functions*

$$\begin{aligned} f'(y_1, y_2, x) &= (y_1 \oplus y_2)\left(f(x) \oplus \widetilde{f}(x)\right) \oplus f(x) \oplus y_1 y_2, \\ f''(y_1, y_2, x) &= (y_1 \oplus y_2)\,(\varphi(x) \oplus \omega(x)) \oplus \varphi(x) \oplus \alpha_1 y_1 \oplus \alpha_2 y_2 \oplus y_1 y_2, \end{aligned}$$

where

$$\begin{gathered} y_1, y_2, \alpha_1, \alpha_2 \in \mathbb{F}_2, \alpha_1 \oplus \alpha_2 = 1, x \in \mathbb{F}_2^n, \\ f \in \mathcal{B}_n, \varphi \in \mathrm{SB}^+(n), \omega \in \mathrm{SB}^-(n), \end{gathered}$$

are self-dual bent functions in $n+2$ variables.

The first construction (for f') was earlier presented in [4] as an example of the construction which uses the indirect sum of bent functions, see [3]. It is worth noting that the second construction (for f'') can also be obtained from indirect sum of bent functions.

Since these constructions do not intersect, the sum of their cardinalities provides a lower bound for the cardinality of the set of self-dual bent iterative functions [19]:

Corollary 4. *It holds*

$$|\mathcal{B}_{n-2}| + \left|\mathrm{SB}^+(n-2)\right|^2 \leqslant \left|\mathrm{SB}^+_{\mathcal{BI}}(n)\right| \leqslant |\mathcal{B}_{n-2}|^2.$$

4.3 The dimension of linear span of sign functions of self-dual bent functions

Let $H_n = H_1^{\otimes n}$ be the n-fold tensor product of the matrix H_1 with itself, where

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

It is known the Hadamard property of this matrix:

$$H_n H_n^{\mathrm{T}} = 2^n I_{2^n}.$$

Denote $\mathcal{H}_n = 2^{-n/2} H_n$. In terms of sign functions the function $f \in \mathcal{F}_n$ is bent if for its sign function F it holds $\mathcal{H}_n F \in \{\pm 1\}^{2^n}$.

Recall that a non-zero vector $v \in \mathbb{C}^n$ is called an *eigenvector* of a square $n \times n$ matrix A attached to the eigenvalue $\lambda \in \mathbb{C}$ if $Av = \lambda v$. A linear span of eigenvectors attached to the eigenvalue λ is called an *eigenspace* associated with λ. Consider a linear mapping $\psi : \mathbb{C}^n \to \mathbb{C}^n$ represented by a $n \times n$ complex matrix A. A *kernel* of ψ is the set

$$\mathrm{Ker}\,(\psi) = \{x \in \mathbb{C}^n : Ax = \mathbf{0} \in \mathbb{C}^n\},$$

where $\mathbf{0}$ is a zero element of the space $\mathbb{C}^n$.

From the definition of self-duality it follows that sign function of any self-dual bent function is the eigenvector of $\mathcal{H}_n$ attached to the eigenvalue 1, that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n}) = \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$. The same holds for a sign function of any anti-self-dual bent function, which obviously is an eigenvector of $\mathcal{H}_n$ attached to the eigenvalue (-1), that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}) = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$.

In [4], an orthogonal decomposition of $\mathbb{R}^{2^n}$ in eigenspaces of H_n was given:

$$\mathbb{R}^{2^n} = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right) \oplus \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right), \tag{2}$$

where the symbol $\oplus$ denotes a direct sum of subspaces.

It is known that

$$\dim\left(\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)\right) = \dim\left(\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)\right) = 2^{n-1},$$

where $\dim(V)$ is the dimension of the subspace $V \subseteq \mathbb{R}^{2^n}$. Moreover, from symmetricity of $\mathcal{H}_n$ it follows that the subspaces $\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$ and $\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$ are mutually orthogonal.

In [19] it was proved that

Theorem 9 ([19])**.** *If $n \geqslant 4$, then:*

- *among sign functions of self-dual bent functions in n variables there exists a basis of the eigenspace of the matrix H_n attached to the eigenvalues 1, that is the subspace $Ker\left(H_n - 2^{n/2} I_{2^n}\right)$;*
- *among sign functions of anti-self-dual bent functions in n variables there exists a basis of the eigenspace of the matrix H_n attached to the eigenvalues (-1), that is the subspace $Ker\left(H_n + 2^{n/2} I_{2^n}\right)$.*

It is worth notice that there exists an example of basis which consists of sign functions of self-dual bent iterative functions provided by two constructions of self-dual bent iterative functions obtained by Theorem 8. Given the basis for self-dual case, the basis for anti-self-dual case can be obtained by using one of bijections from Theorem 20.

5 Metrical complement and regularity

In this section, we give results regarding notable metrical property of a subset of Boolean cube called metrical regularity. The sets of affine Boolean functions and bent functions possess it. The sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables are also mutually maximally distant. That implies metrical *duality*, in some sence, between the considered pairs of subsets of Boolean functions.

Regarding that, some essential and intriguing questions arise: for instance, are there any pairs of metrically regular subsets inside the metrically regular set of bent functions in n variables? If additionally, in order to exclude some trivial cases, we consider only the subsets which include functions together with their negations, the maximal Hamming distance from the considered sets is at most 2^{n-1}. Are there any pairs of metrically regular subsets with additional mentioned requirement such that the distance between them is exactly 2^{n-1}, that is to say they are extremal in a manner?

5.1 Definitions

Let $X \subseteq \mathbb{F}_2^n$ be an arbitrary set and let $y \in \mathbb{F}_2^n$ be an arbitrary vector. Define the *distance* between y and X as $\mathrm{dist}(y, X) = \min_{x \in X} \mathrm{dist}(y, x)$. The *maximal distance* from the set X is

$$\mathrm{d}(X) = \max_{y \in \mathbb{F}_2^n} \mathrm{dist}(y, X).$$

In coding theory this number is also known as the *covering radius* of the set X. A vector $z \in \mathbb{F}_2^n$ is called *maximally distant* from a set X if $\mathrm{dist}(z, X) = \mathrm{d}(X)$. The set of all maximally distant vectors from the set X is called the *metrical complement* of the set X and denoted by $\widehat{X}$. A set X is said to be *metrically regular* if $\widehat{\widehat{X}} = X$. Define, following N. Tokareva [39], a subset of Boolean functions to be *metrically regular* if the set of corresponding vectors of values is metrically regular.

Sets of functions which have maximum distance from partition set functions were studied in [33], it was shown that partition set functions defined by some partition are mutually maximally distant sets. Lower bound on size of the largest metrically regular subset of the Boolean cube was studied in [28].

5.2 The set of bent functions

Let $\mathrm{GA}(n)$ stands for the affine group. It is well-known that

Proposition 1. *Any isometric mapping of the form*

$$f(x) \longrightarrow f\left(Ax \oplus b\right) \oplus \langle c, x \rangle \oplus d,$$

where $A \in \mathrm{GL}(n)$, $b, c \in \mathbb{F}_2^n$, $d \in \mathbb{F}_2$, preserves bentness.

In [36] the following theorem was proved:

Theorem 10 ([36])**.** *For each non-affine Boolean function $h \in \mathcal{F}_n$ there exists a bent function $f \in \mathcal{B}_n$ such that $f \oplus h$ is not bent.*

From Proposition 1 and Theorem 10 it follows that the set of bent functions is closed under addition of affine Boolean functions only. This fact implies that the affine functions are precisely all Boolean functions which are at the maximum distance from the class of bent functions. Namely, in [36] it was shown that

Theorem 11 ([36])**.** *A Boolean function in n variables is*

- *a bent function if and only if it has the maximal possible distance $2^{n-1} - 2^{n/2-1}$ to the set of all affine functions, that is it is an element of $\widehat{\mathcal{A}_n}$;*
- *an affine function if and only if it has the maximal possible distance $2^{n-1} - 2^{n/2-1}$ to the set of all bent functions, that is it is an element of $\widehat{\mathcal{B}_n}$.*

Thus, from the results given in [36] it follows that there exists a *duality*, in some sense, between the definitions of bent functions and affine functions. In particular, we obtain metrical regularity of the sets of affine functions and bent functions.

Corollary 5.

1. *The set $\mathcal{A}_n$ of all affine Boolean functions in n variables is metrically regular.*
2. *The set $\mathcal{B}_n$ of all bent functions in n variables is metrically regular.*

5.3 The set of (anti-)self-dual bent functions

For any (anti-)self-dual bent function $f \in \mathrm{SB}^+(n)$ its negation $f \oplus 1$ is also (anti-)self-dual bent [4, 12]. Moreover, from the results presented in [20], it follows the counterpart of Theorem 10 for the (anti-)self-dual case, namely:

Theorem 12. *For each non-constant Boolean function $h \in \mathcal{F}_n$ there exists a self-dual bent function $f \in \mathrm{SB}^+(n)$ such that $f \oplus h$ is not self-dual bent. Anti-self-dual bent functions possess the same property.*

So, it follows that the set of (anti-)self-dual bent functions is closed only under addition of 1, that is, taking the negation of the function.

From the fact that considered set is closed under addition of 1, it follows that the maximal Hamming distance from the set $\mathrm{SB}^+(n)$ is at most 2^{n-1}. It was proved by Carlet et al. in [4] that the Hamming distance between any pair of self-dual and anti-self-dual bent functions, both in n variables, is equal to 2^{n-1}. So, we have

$$\mathrm{d}\left(\mathrm{SB}^+(n)\right) = 2^{n-1},$$

and all anti-self-dual bent functions in n variables belong to the metrical complement of the set of self-dual bent functions in n variables.

In paper [19], the metrical complement of the set of (anti-)self-dual bent functions in $n \geqslant 4$ variables was completely characterized by using the orthogonal decomposition (2) and existence of the basis provided by the Theorem 9, namely, it was proven that

Theorem 13 ([19])**.** *Let $n \geqslant 4$, then a Boolean function in n variables is:*

- *self-dual bent if and only if it has the maximal possible distance 2^{n-1} to the set of all anti-self-dual bent functions, that is, it is an element of $\widehat{\mathrm{SB}^-(n)}$;*
- *anti-self-dual bent if and only if it has the maximal possible distance 2^{n-1} to the set of all self-dual bent functions, that is, it is an element of $\widehat{\mathrm{SB}^+(n)}$.*

As for the pair of the sets of bent functions and affine functions, it follows that there also exists a *duality* between the sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables.

The case $n = 2$ was considered explicitely and it appeared that both $\mathrm{SB}^+(2)$ and $\mathrm{SB}^-(2)$ are metrically regular sets. From that and the Theorem 13 it follows

Corollary 6.

1. *The set $\mathrm{SB}^+(n)$ of all self-dual bent functions in n variables is metrically regular.*
2. *The set $\mathrm{SB}^-(n)$ of all anti-self-dual bent functions in n variables is metrically regular.*

6 The group of automorphisms

Study of automorphism groups of mathematical objects deserves attention since these groups are closely connected with the structure of the objects. There exists a natural question: how groups of automorphisms of two mathematical objects, one of which is embedded to another one, are related.

An example of such a problem statement is the set of bent functions in n variables and one of its significant subclasses which consisits of self-dual bent functions in n variables.

It is also worth mentioning that the complexity of classification of combinatorial objects depends on generality of the approach. Consequently, the question "*if the common approach to classify (self-dual) bent functions is the most general within automorphisms of the set of Boolean functions*", arises naturally.

6.1 Isometric mappings and automorphism groups

A mapping φ of the set of all Boolean functions in n variables to itself is called *isometric* if it preserves the Hamming distance between functions, that is,

$$\operatorname{dist}(\varphi(f), \varphi(g)) = \operatorname{dist}(f, g)$$

for any $f, g \in \mathcal{F}_n$. Following [20], denote the set of all isometric mappings of the set of all Boolean functions in n variables to itself by $\mathcal{I}_n$.

It is known (A. A. Markov, 1956) that every isometric mapping of all Boolean functions in n variables to itself has the unique representation of the form

$$f(x) \longrightarrow f(\pi(x)) \oplus g(x), \tag{3}$$

where π is a permutation on the set $\mathbb{F}_2^n$ and $g \in \mathcal{F}_n$ [23]. The mapping of this form is denoted by $\varphi_{\pi,g} \in \mathcal{I}_n$.

The *group of automorphisms* of a fixed subset $M \subseteq \mathcal{F}_n$ is the group of isometric mappings of the set of all Boolean functions in n variables to itself preserving the set M. It is denoted by $\operatorname{Aut}(M)$.

6.2 Matrix representation

For a number $k \in \{0, 1, \ldots, 2^n - 1\}$ denote by $\mathbf{v}_k \in \mathbb{F}_2^n$ its binary representation.

Recall that a square matrix is called *monomial* (or *generalized permutation matrix*) if it has exactly one nonzero entry in each row and each column.

The following one-to-one correspondence between the set $\mathcal{I}_n$ and the set of monomial matrices of order $2^n \times 2^n$ with nonzero elements from the set $\{\pm 1\}$ was used in [20]. In more detail, let $\varphi_{\pi,g} \in \mathcal{I}_n$ be an arbitrary isometric mapping. Then for any $f \in \mathcal{F}_n$ and its sign function

$$F = \left((-1)^{f(\mathbf{v}_0)}, (-1)^{f(\mathbf{v}_1)}, \ldots, (-1)^{f(\mathbf{v}_{2^n-1})}\right) \in \{\pm 1\}^{2^n},$$

the sign function

$$F' = \left((-1)^{f'(\mathbf{v}_0)}, (-1)^{f'(\mathbf{v}_1)}, \ldots, (-1)^{f'(\mathbf{v}_{2^n-1})}\right) \in \{\pm 1\}^{2^n},$$

of $f' = \varphi_{\pi,g}(f) \in \mathcal{F}_n$ can be expressed as $F' = AF$, where A is a $2^n \times 2^n$ monomial matrix, constructed by the permutation π and the function g:

$$i \begin{pmatrix} & & & j & & & \\ & & & \vdots & & & \\ & & & 0 & & & \\ & & & \vdots & & & \\ \ldots & 0 & \ldots & (-1)^{g(\mathbf{v}_{i-1})} & \ldots & 0 & \ldots \\ & & & \vdots & & & \\ & & & 0 & & & \\ & & & \vdots & & & \end{pmatrix},$$

in which in the i-th row a nonzero element $(-1)^{g(\mathbf{v}_{i-1})}$ is in the j-th column, where $(j-1)$ is a number with binary representation $\pi(\mathbf{v}_{i-1})$. So the i-th component of $F' = AF$ is equal to

$$(-1)^{f'(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))}(-1)^{g(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))\oplus g(\mathbf{v}_{i-1})}$$

for any $i \in \{1, 2, \ldots, 2^n\}$, that is equivalent to

$$f'(x) = f(\pi(x)) \oplus g(x), \quad x \in \mathbb{F}_2^n.$$

6.3 The group of automorphisms of the set of bent functions

Some attempts to determine the automorphism group of a given bent function were undertaken by U. Dempwolff in 2006 [10]. Results were presented in terms of elementary Abelian Hadamard difference sets (equivalently, bent functions).

A natural question whether there exist isometric mappings of Boolean functions into itself, distinct from those mentioned in Proposition 1, which preserve the class of bent function, was completely solved in paper [34]. It was proved that there were no other mappings possessing such a property. Namely, by using the Theorem 11 in view of the duality, the following coincidence was shown.

Theorem 14 ([34]). $\mathrm{Aut}\,(\mathcal{B}_n) = \mathrm{Aut}\,(\mathcal{A}_n)$.

The group of automorphisms of the set of all affine functions in n variables consists, as it is well known, of mappings of the form (3) with affine permutation π and affine shift g, see, for example, [22]. Note that the set of all affine functions in n variables forms a group isomorphic to $\mathbb{F}_2^{n+1}$. Let the symbol $\ltimes$ stands for the semidirect product, then the result is formulated as follows.

Theorem 15 ([34]). $\mathrm{Aut}\,(\mathcal{B}_n) = \mathrm{GA}(n) \ltimes \mathbb{F}_2^{n+1}$.

These results imply the non-existence of a more general approach to equivalence of bent functions than that on the base of isometric mappings.

6.4 The group of automorphisms of the set of (anti-)self-dual bent functions

In [4] the following problem was pointed:

Open question (Carlet, Danielson, Parker, Solé, 2010): to find mappings preserving self-duality, distinct from the known ones, or give a proof that there are no more.

In [20], this question was resolved within isometric mappings of the set of all Boolean functions in $n \geqslant 4$ variables into itself.

At first there is the problem of how the sets of isometric mapping preserving self-duality and anti-self-duality or, in other words, groups of automorphisms of the sets $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ are related. This problem was solved in [20], where with a use of the orthogonal decomposition (2) and the basis from the Theorem 9 it was proved that

Theorem 16 ([20]). *If $n \geqslant 4$, then* $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right)$.

In [20] the criterion of preserving self-duality was also presented:

Theorem 17 ([20]). *If $n \geqslant 4$, then isometric mapping $\varphi_{\pi,g}$ belongs to* $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right)$ *if and only if, for any $x, y \in \mathbb{F}_2^n$, it holds*

$$\langle \pi(x), y\rangle \oplus g(x) = \left\langle x, \pi^{-1}(y)\right\rangle \oplus g\left(\pi^{-1}(y)\right).$$

In matrix terms the criterion can be formulated as $A\mathcal{H}_n = \mathcal{H}_n A$, where A is the matrix which represents the mapping $\varphi_{\pi,g}$.

The problem of characterization mappings which preserve self-duality was studied in [4, 12], where it was shown that the mapping (1) preserves self-duality of a bent function, in other words, it is an element of $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right)$. It is obvious that this mapping is isometric and corresponds to $\varphi_{\pi,g} \in \mathcal{I}_n$ with

$$\pi(x) = L\,(x \oplus c)\,,\ g(x) = \langle c, x\rangle \oplus d,\ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$. The group which consists of mappings of such form is called an *extended orthogonal group* and denoted by $\overline{\mathcal{O}}_n$ [9, 12]. It is known that this group is a subgroup of $\mathrm{GL}\,(n+2, \mathbb{F}_2)$ [12].

In paper [20] known results were generalized within isometric mappings from the set $\mathcal{I}_n$ for $n \geqslant 4$. Namely, by using the criterion from Theorem 17 and the matrix representation of isometric mappings (see Section 6.2) it was proved that the desired group of automorphisms coincides with the extended orthogonal group.

Theorem 18 ([20]). *For $n \geqslant 4$ it holds*

$$\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right) = \overline{\mathcal{O}}_n.$$

It follows that the classification of self-dual bent functions in $n \geqslant 4$ variables based on the restricted form of affine equivalence proposed in articles [4, 12] is the most general within isometric mappings of the set of all Boolean functions in n variables into itself.

7 Isometric mappings and duality

In this section we discuss results from paper [20] on characterization of isometric mappings which define bijections between self-dual and anti-self dual bent functions, and description of isometric mappings which preserve or change the sign of the Rayleigh quotient of a Boolean function.

7.1 Isometric bijections between self-dual and anti-self-dual bent functions

It is known [4] that there exists a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$, based on the decomposition of sign functions of (anti-)self-dual bent functions. Also note that from the existence of such bijection it follows that $\left|\mathrm{SB}^+(n)\right| = \left|\mathrm{SB}^-(n)\right|$.

Namely, let $(Y, Z) \in \{\pm 1\}^{2^n}$, where $Y, Z \in \{\pm 1\}^{2^{n-1}}$, be a sign function for some $f \in \mathrm{SB}^+(n)$. Then a vector $(Z, -Y) \in \{\pm 1\}^{2^n}$ is a sign function for some function from $\mathrm{SB}^-(n)$. In terms of isometric mappings the mentioned transformation can be represented as

$$f(x) \longrightarrow f(x \oplus c) \oplus \langle c, x \rangle,$$

where $c = (1, 0, 0, ..., 0) \in \mathbb{F}_2^n$.

In paper [15] it was mentioned that the more general form of this mapping

$$f(x) \longrightarrow f(x \oplus c) \oplus \langle c, x \rangle,$$

where $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$. It is obvious that this mapping is an element from $\mathcal{I}_n$.

In paper [20] these results were generalized within isometric mappings from the set $\mathcal{I}_n$ for $n \geqslant 4$.

The criterion of bijectivity between self-dual and anti-self-dual bent functions was obtained in [20] with a use of the orthogonal decomposition (2) and the basis from the Theorem 9.

Theorem 19 ([20]). *Let $n \geqslant 4$, then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if, for any $x, y \in \mathbb{F}_2^n$, it holds*

$$\langle \pi(x), y \rangle \oplus g(x) = \left\langle x, \pi^{-1}(y) \right\rangle \oplus g\left(\pi^{-1}(y)\right) \oplus 1.$$

By using this criterion in [20] the general form of considered isometric bijections was found.

Theorem 20 ([20]). *For $n \geqslant 4$ isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if*

$$\pi(x) = L(x \oplus c), \ g(x) = \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, $d \in \mathbb{F}_2$.

Thus, from Theorems 18 and 20 we can conclude that if we take a mapping from the group $\overline{\mathcal{O}}_n$ and replace the vector $c \in \mathbb{F}_2^n$ by a binary vector of length n with an odd Hamming weight then we switch the mapping from the "automorphism mode" to the "bijection mode" between the sets $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$.

7.2 Isometric mappings and the Rayleigh quotient

In [4] the *Rayleigh quotient* S_f of a Boolean function $f \in \mathcal{F}_n$ was defined as

$$S_f = \sum_{x,y \in \mathbb{F}_2^n} (-1)^{f(x) \oplus f(y) \oplus \langle x,y \rangle} = \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)} W_f(y).$$

In a scope of bent functions the Rayleigh quotient characterizes the Hamming distance between a bent function and its dual. Indeed, let $f \in \mathcal{B}_n$, then

$$\operatorname{dist}(f, \widetilde{f}) = 2^{n-1} - \frac{1}{2^{n/2+1}} S_f = 2^{n-1} - \frac{1}{2} N_f.$$

In [4] it was proved that for any $f \in \mathcal{F}_n$ the absolute value of S_f is at most $2^{3n/2}$ with equality if and only if f is self-dual $\left(+2^{3n/2}\right)$ and anti-self-dual $\left(-2^{3n/2}\right)$ bent function. That is the maximum (minimum) value of the Rayleigh quotient of a Boolean function in an even number of variables is attainable on self-dual (anti-self-dual) bent functions and only them, thus providing a criterion for (anti-)self-duality in terms of the Rayleigh quotient values.

In article [9] the operations on Boolean functions that preserve bentness and the Rayleigh quotient were given. Namely, it was proved that for any $f \in \mathcal{B}_n, L \in \mathcal{O}_n, c \in \mathbb{F}_2^n, d \in \mathbb{F}_2$ the functions $g, h \in \mathcal{B}_n$ defined as $g(x) = f(Lx) \oplus d$ and $h(x) = f(x \oplus c) \oplus \langle c, x \rangle$ provide $N_g = N_f$ and $N_h = (-1)^{\langle c,c \rangle} N_f$.

The mentioned operations are isometric mappings from $\mathcal{I}_n$. The complete characterization of isometric mappings that preserve the Rayleigh quotient as well as change it was given in [20].

Theorem 21 ([20]). *If $n \geqslant 4$ then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ preserves the Rayleigh quotient of every Boolean function in n variables if and only if $\varphi_{\pi,g} \in \operatorname{Aut}\left(\mathrm{SB}^+(n)\right)$.*

Theorem 22 ([20]). *If $n \geqslant 4$ then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ changes the sign of the Rayleigh quotient of every Boolean function in n variables if and only if it is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$.*

In a scope of bent functions the Rayleigh quotient characterizes the Hamming distance between a bent function and its dual. Indeed, let $f \in \mathcal{B}_n$, then

$$\operatorname{dist}(f, \widetilde{f}) = 2^{n-1} - \frac{S_f}{2^{n/2+1}}.$$

So, from Theorem 21 we immediately have that general form of isometric mappings, which preserve the Hamming distance between every bent function and its dual, is described by the extended orthogonal group $\overline{\mathcal{O}}_n$ (see Theorem 2).

Conclusion

In this work, we have given a review of metrical properties of the set of bent functions and its subset of functions which coincide with their duals. The group of automorphisms and metrical complements of these sets are described. We also reviewed some general metrical properties of the set of self-dual bent functions and considered an iterative construction of bent functions. Some relevant open problems and hypothesis on bent functions weere discussed.

An interesting question is the characterization of isometric mappings preserving bentness and self-duality, that are beyond the automorphisms of the set of all Boolean functions.

The solution of the problems, that were considered in this review, with regard to different generalizations of bent functions that is study of metrical properties and the duality as well as self-duality in this scope is a goal worth pursuing.

References

[1] Canteaut A., Charpin P., Decomposing bent functions, IEEE Trans. Inform. Theory, **49**(8), 2004–2019 (2003).

[2] Canteaut A., Daum M., Dobertin H., Leander G., Finding nonnormal bent functions, Discrete Appl. Math., **154**(2), 202–218 (2006).

[3] Carlet C. Boolean functions for cryptography and error correcting code. In: Crama Y., Hammer P.L. (eds.) Boolean Models and Methods in Mathematics, Computer Science, and Engineering. p. 257–397. Cambridge University Press, Cambridge (2010).

[4] Carlet C., Danielson L.E., Parker M.G., Solé P., Self-dual bent functions, Int. J. Inform. Coding Theory, **1**, 384–399 (2010).

[5] Carlet C., Mesnager S., Four decades of research on bent functions, Des. Codes Cryptogr., **78**(1), 5–50 (2016).

[6] Climent J.-J., Garcia F.J., Requena V., A construction of bent functions of $n+2$ variables from a bent function of n variables and its cyclic shifts, Algebra. https://doi.org/10.1155/2014/701298 (2014).

[7] Coulter R., Mesnager S. Bent functions from involutions over $\mathbb{F}_{2^n}$, IEEE Trans. Inf. Theory, **64**(4), 2979–2986 (2018).

[8] Cusick T.W., Stănică P., Cryptographic Boolean functions and applications, Acad. Press, London, 2017, 288 pp.

[9] Danielsen L.E., Parker M.G., Solé P., The Rayleigh quotient of bent functions, Springer Lect. Notes in Comp. Sci. 5921, pp. 418–432. Springer, Berlin (2009).

[10] Dempwolff U., Automorphisms and Equivalence of Bent Functions and of Difference Sets in Elementary Abelian 2-Groups, Commun. Algebra, **34**(3), 1077–1131 (2006).

[11] Dillon J., Elementary Hadamard Difference Sets, PhD. dissertation, Univ. Maryland, College Park (1974).

[12] Feulner T., Sok L., Solé P., Wassermann A. Towards the Classification of Self-Dual Bent Functions in Eight Variables. Des. Codes Cryptogr. **68**(1), 395–406 (2013).

[13] Janusz G.J., Parametrization of self-dual codes by orthogonal matrices, Finite Fields Appl., **13**(3), 450–491 (2007).

[14] Hou X.-D., New Constructions of Bent Functions, Proc. of International Conference on Combinatorics, Information Theory and Statistics. Journal of Combinatorics, Information and System Sciences, **25**(1–4), 173–189 (2000).

[15] Hou X.-D., Classification of self dual quadratic bent functions, Des. Codes Cryptogr. **63**(2), 183–198 (2012).

[16] Hyun J.Y., Lee H., Lee Y., MacWilliams duality and Gleason-type theorem on self-dual bent functions, Des. Codes Cryptogr., **63**(3), 295–304 (2012).

[17] Kolomeec N.A., The Graph of Minimal Distances of Bent Functions and Its Properties, Des. Codes Cryptogr., **85**(3), 1–16 (2017).

[18] Kutsenko A.V., The Hamming Distance Spectrum Between Self-Dual Maiorana–McFarland Bent Functions, Journal of Applied and Industrial Mathematics, **12**(1), 112–125 (2018).

[19] Kutsenko A., Metrical properties of self-dual bent functions, Des. Codes Cryptogr., **88**(1), 201–222 (2020).

[20] Kutsenko A., The group of automorphisms of the set of self-dual bent functions, Cryptogr. Commun., 2020, (submitted).

[21] Luo G., Cao X., Mesnager S. Several new classes of self-dual bent functions derived from involutions, Cryptogr. Commun., **11**(6), 1261–1273 (2019).

[22] MacWilliams F.J., Sloane N.J.A, The Theory of Error Correcting Codes, North-Holland Publishing Company, (1977).

[23] Markov A. A. On transformations without error propagation, in: Selected Works, Vol. II: Theory of Algorithms and Constructive Mathematics. Mathematical Logic. Informatics and Related Topics, p. 70–93, MTsNMO, Moscow (2003) [in Russian].

[24] McFarland R.L., A family of difference sets in non-cyclic groups, J. Combin. Theory Ser. A, **15**(1), 1–10 (1973).

[25] Mesnager S., Several New Infinite Families of Bent Functions and Their Duals, IEEE Trans. Inf. Theory, **60**(7), 4397–4407 (2014).

[26] Mesnager S., Bent Functions: Fundamentals and Results. Springer, Switzerland (2016).

[27] Mesnager S. On constructions of bent functions from involutions, In: Proceedings of ISIT, pp. 110–114 (2016).

[28] Oblaukhov A.K. A lower bound on the size of the largest metrically regular subset of the Boolean cube. Cryptogr. Commun., **11**(4), 777–791 (2019).

[29] Preneel B., Van Leekwijck W., Van Linden L., Govaerts R., Vandewalle J., Propagation characteristics of Boolean functions. In: Advances in Cryptology-EUROCRYPT. Lecture Notes in Computer Science, vol. 473, pp. 161–173. Springer, Berlin (1990).

[30] Rifà J., Zinoviev V. A. On binary quadratic symmetric bent and almost bent functions. arXiv:1211.5257v3 (2019).

[31] Rothaus O.S., On bent functions, J. Combin. Theory. Ser. A, **20**(3), 300–305 (1976).

[32] Sok L., Shi M., Solé. P., Classification and Construction of quaternary self-dual bent functions, Cryptogr. Commun., **10**(2), 277–289 (2018).

[33] Stănică P., Sasao T., Butler J.T., Distance duality on some classes of Boolean functions, J. Combin. Math. and Combin. Computing, **107**, 181–198 (2018).

[34] Tokareva N.N., The group of automorphisms of the set of bent functions, Discrete Mathematics and Applications, **20**(5), 655–664 (2010).

[35] Tokareva N.N., On the number of bent functions from iterative constructions: lower bounds, Adv. Math. Commun., **5**(4), 609–621 (2011).

[36] Tokareva N., Duality between bent functions and affine functions, Discrete Math., 2012, **312**(3), 666–670 (2012).

[37] Tokareva N.N., On decomposition of a Boolean function into sum of bent functions, Siberian Electronic Mathematical Reports, **11**, 745–751 (2014).

[38] Tokareva N.N., On Decomposition of a Dual Bent Function into Sum of Two Bent Functions, Prikl. Diskretn. Mat., **26**(4), 59–61 (2014). [in Russian]

[39] Tokareva N., Bent Functions, Results and Applications to Cryptography, 230 p., Acad. Press. Elsevier (2015).

UDC 519.7 DOI 10.17223/20710410/XX/1

ON METRIC COMPLEMENTS AND METRIC REGULARITY IN FINITE METRIC SPACES [1]

A. K. Oblaukhov

Sobolev Institute of Mathematics, Novosibirsk, Russia,
Novosibirsk State University, Novosibirsk, Russia,
Laboratory of Cryptography JetBrains Research, Novosibirsk, Russia

E-mail: oblaukhov@gmail.com

This review is devoted to the study of metric complements and metric regularity in the Boolean cube and in arbitrary finite metric spaces. Let A be an arbitrary subset a finite metric space M, and $\widehat{A}$ be the *metric complement* of A — the set of all points of M at the maximal possible distance from A. If the metric complement of the set $\widehat{A}$ coincides with A, then the set A is called a *metrically regular set*. The problem of investigating metrically regular sets was posed by N. Tokareva in 2012 when studying metric properties of *bent functions*, which have important applications in cryptography and coding theory and are also one of the earliest examples of a metrically regular set. In this paper, main known problems and results concerning the topic of metric regularity are overviewed, such as the problem of finding the largest and the smallest metrically regular set, both in the general case and in the case of fixed covering radius, and the problem of obtaining metric complements and establishing metric regularity of linear codes. Results concerning metric regularity of partition sets of functions and Reed-Muller codes are presented.

Keywords: *metrically regular set, metric complement, covering radius, bent function, deep hole, Reed-Muller code, linear code*

Данный обзор посвящён исследованиям метрических дополнений и метрической регулярности в булевом кубе и в произвольных конечных метрических пространствах. Пусть A — произвольное подмножество конечного метрического пространства M, а $\widehat{A}$ — метрическое дополнение A — множество всех точек M, находящихся на максимально возможном расстоянии от A. Если метрическое дополнение множества $\widehat{A}$ совпадает с множеством A, то A называется *метрически регулярным*. Задача изучения метрически регулярных множеств была поставлена Н. Токаревой в 2012 году в процессе изучения метрических свойств *бент-функций*. Бент-функции имеют важные приложения в криптографии и теории кодирования, а также являются одним из первых примеров метрически регулярного множества. В данной работе проводится обзор основных задач и результатов, связанных с понятием метрической регулярности, в частности, задача поиска наибольших и наименьших метрически регулярных множеств (как в общем случае, так и в случае фиксированного радиуса покрытия), задача нахождения метрического дополнения и определения метрической регулярности линейных кодов. Приведены результаты, связанные с метрической регулярностью множеств функций, построенных на разбиении пространства, а также кодов Рида-Маллера.

[1]The work was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by Russian Foundation for Basic Research (projects no. 18-07-01394, 19-31-90093) and Laboratory of Cryptography JetBrains Research.

Ключевые слова: *метрически регулярное множество, метрическое дополнение, радиус покрытия, бент-функция, код Рида-Маллера, линейный код*

Introduction

The problem of investigating and classifying *metrically regular sets* was posed by Tokareva [24,25] when studying metric properties of *bent functions* [20]. A Boolean function f in even number of variables m is called a *bent function* if it is at the maximal possible distance from the set of *affine functions*.

Bent functions have various applications in cryptography, coding theory and combinatorics [4, 13, 25]. In cryptography, bent functions are valued because of their outstanding nonlinearity, which helps to construct S-boxes for block ciphers with high resistance to linear cryptanalysis, and, as it turned out, good diffusion properties and high resistance to differential cryptanalysis [13]. Bent functions were also used in the construction of the stream cipher Grain, being a part of a nonlinear feedback shift register [25]. From the coding theory standpoint, bent functions form the set of points at the maximal possible distance from the Reed-Muller code of the first order $\mathcal{RM}(1,m)$ in even number of variables m. Bent functions are used to construct Kerdock codes, which are optimal and have large code distances (see more in [13]). Bent functions also have a number of representations and relations to different combinatorial objects: Hadamard difference sets, block designs, etc. (see [13,25]).

However, many problems related to bent functions remain unsolved; in particular, the gap between the best known lower and upper bound on the number of bent functions is extremely large; currently known constructions of bent functions are rather scarse.

In 2010 [23], Tokareva has proved that, like bent functions are maximally distant from affine functions, affine functions are at the maximal possible distance from bent functions, thus establishing the *metric regularity* of both sets. Combined with the importance of bent functions in cryptography and coding theory, this arouses the interest in studying the property of metric regularity and in the classification of metrically regular sets.

This paper is devoted to the study of metrically regular sets, both in the Boolean cube and in arbitrary finite metric spaces. Published results concerning the topic, as well as some currently unpublished, are overviewed.

The first section provides necessary basic definitions, simple examples of metrically regular sets and some of their trivial properties. Section 2 describes the results of Stănică, Sasao and Butler [22] concerning metric complements and metric regularity of *partition sets of functions*. Section 3 deals with the problem of finding the smallest and the largest metrically regular sets, both in general and in the case of fixed distance between sets [17]. *Strongly metrically regular sets* are introduced in Section 4 as a subclass of metrically regular sets. These allow one to obtain iterative constructions of metrically regular sets and get an estimate on how big the largest metrically regular set with fixed covering radius can be [18]. Section 5 touches upon the problem of describing metric complements and establishing metric regularity of linear codes. General results are presented, and the metric regularity of several families of Reed-Muller codes is established [16, 19].

1. Preliminaries

1.1. Definitions

Let M be a finite discrete metric space with a metric $d(\cdot,\cdot)$, which admits values from a set D. From now on every space mentioned in the paper will be a finite discrete metric space. Let $X \subseteq M$ be an arbitrary subset of the space and $y \in M$ be an arbitrary point.

The distance $d(y, X)$ from the point y to the set X is equal to $\min\limits_{x \in X} d(y, x)$. The *covering radius* of the set X is defined as follows:

$$\rho(X) = \max_{z \in M} d(z, X).$$

A set X with the covering radius r is also sometimes called a *covering code* [3] of radius r.

Consider the following set

$$\{y \in M : d(y, X) = \rho(X)\}$$

of all vectors at the maximal possible distance from the set X. This set is called the *metric complement* [16] of X and is denoted by $\widehat{X}$. If $\widehat{\widehat{X}} = X$, the set X is said to be *metrically regular* [24].

Note that metrically regular sets always come in pairs, i.e. if A is a metrically regular set, its metric complement $\widehat{A}$ is also a metrically regular set. In this work a pair consisting of a metrically regular set A and its metric complement $B = \widehat{A}$ will sometimes be referred to as "a pair of metrically regular sets A, B".

Throughout the paper we will mostly consider the metric space $\mathbb{F}_2^n$ of binary vectors of length n equipped with the Hamming metric. The *Hamming distance* $d_H(\cdot, \cdot)$ between two binary vectors is defined as the number of coordinates in which these vectors differ, while $wt(\cdot)$ denotes the *Hamming weight* of a vector, i.e. the number of nonzero values it contains. Since $\mathbb{F}_2$ is a field, $\mathbb{F}_2^n$ is also considered as a vector space, with the plus sign $+$ denoting addition of vectors modulo two. A *Boolean function* in m variables is an arbitrary mapping from $\mathbb{F}_2^m$ to $\mathbb{F}_2$.

1.2. Examples and basic results

Let us consider some simple examples of metric complements and metrically regular sets in the space $\mathbb{F}_2^n$:

1) Let $X = \{x\}$ be the set consisting of one binary vector. It has covering radius n and its metric complement is the set $\widehat{X} = \{x + \mathbf{1}\}$, consisting only of the opposite vector (here $\mathbf{1}$ is the all-ones vector). It follows that $\widehat{\widehat{X}} = X$, so X is a metrically regular set;
2) Consider a ball of radius r centered at x; i.e., $X = \{y \in \mathbb{F}_2^n | d(x, y) \leqslant r\}$. Then the vector $x + \mathbf{1}$ will be at the distance $n - r$ from the set X, while any other vector will be at a smaller distance. Therefore, the covering radius of X is equal to $n - r$ and its metric complement is the set $\widehat{X} = \{x + \mathbf{1}\}$. Then $\widehat{\widehat{X}} = \{x\}$, which shows us that, unless $r = 0$, the ball of radius r is not a metrically regular set.

For other examples of metric complements and metrically regular sets the reader is referred to [16–18].

Let us return to an arbitrary metric space M with a metric admitting values from a set D and present some basic results concerning metric regularity.

An *automorphism* of a set $X \subseteq M$ is an isometric mapping from M into M which maps X into itself. The following result [16] is straightforward from the definition of metric regularity, and is also described in [23, 24] for affine/bent functions:

Theorem 1. Let $X \subseteq M$ be a metrically regular set. Then sets of automorphisms of X and $\widehat{X}$ coincide: $\mathrm{Aut}(X) = \mathrm{Aut}(\widehat{X})$.

As we could see from examples, not every set is metrically regular, which means that we can apply the procedure of taking metric complement more than twice and obtain new

sets. It has been proven [16] that this process stabilizes for any set after not more than $|D| - 1$ repetitions:

Proposition 1. Let X be an arbitrary subset of the space M. Let us denote $X_0 = X$, $X_{k+1} = \widehat{X}_k$ for $k \geqslant 0$. Then there exists a number $M \leqslant |D| - 1$ such that X_m is a metrically regular set for any $m \geqslant M$.

Using this proposition, we can, for example, split the set 2^M of all subsets of M into equivalence classes, and call two sets X,$Y \subseteq M$ equivalent if and only if the pair of metrically regular sets A, A^*, which we obtain from the set X by repeatedly obtaining metric complement as in Proposition 1, coincides with the pair of metrically regular sets B, B^* which we obtain from the set Y. How would the equivalence classes look? The description has not yet been given.

Proposition 1 is also useful when conducting experiments with metrically regular sets using computers.

2. Partition sets of functions

In the work [22] Stănică, Sasao and Butler introduce the notion of *partition sets of functions* and study their metric complements and metric regularity.

A set $\mathcal{S}$ of Boolean functions is said to be a *partition set* with respect to a partition $\mathcal{U}$ of the set $\mathbb{F}_2^m$, if the elements in the same block of $\mathcal{U}$ all map to 0 or all map to 1, and all combinations of assignments to blocks are included in $\mathcal{S}$. Partition set functions include, for example symmetric functions, rotation symmetric functions, self-anti-dual-functions and linear structure functions.

The following theorem presents the main result of their work, describing the covering radius and the metric complement of a partition set of functions:

Theorem 2. Consider a partition set of functions $\mathcal{S}$, and let us denote the covering radius of $\mathcal{S}$ as $\rho_{\mathcal{S}}$. Let $N_{\mathcal{S}}$ be the number of Boolean functions at distance $\rho_{\mathcal{S}}$ from $\mathcal{S}$. Then,

$$\rho_{\mathcal{S}} = \sum_{i=1}^{l} \lfloor k_i/2 \rfloor \text{ and } N_{\mathcal{S}} = \prod_{i=1}^{l} \frac{1}{2 - k_i \mod 2} \left(\binom{k_i}{\lfloor k_i/2 \rfloor} + \binom{k_i}{\lceil k_i/2 \rceil} \right),$$

where k_i is the cardinality of the i-th block of the l blocks in partition $\mathcal{U}$.

The proof of the theorem is constructive and gives an explicit description of the metric complement $\widehat{\mathcal{S}}$. From this description the equality $\widehat{\widehat{\mathcal{S}}} = \mathcal{S}$ is trivially established, showing that all partition sets of functions are metrically regular.

The authors then proceed to investigate special cases of partition sets of functions, namely, *symmetric* and *rotation symmetric* functions. They calculate covering radii for both of these sets, give characterization for the set of maximally asymmetric functions (the metric complement of the set of symmetric functions) and calculate the number of such functions. They also study the weight distribution of maximally asymmetric functions, as well as their algebraic degrees, and provide a classification of all functions with respect to the distance from the set of symmetric functions. For details, the reader is referred to [22].

3. Largest and smallest metrically regular sets

Let us return to affine and bent functions. Since the gap between the best known upper and lower bounds on the size of the set of bent functions is so large, it is interesting to investigate possible cardinalities of metrically regular sets, particularly, the extreme cardinalities, in an attempt to improve known bounds. The work [17] focuses on the problem of finding the largest and the smallest metrically regular sets.

3.1. General problem

In the Boolean cube $\mathbb{F}_2^n$ with the Hamming distance, any smallest metrically regular set has cardinality 1, as can be seen from the simplest example $X = \{x\}$, $x \in \mathbb{F}_2^n$. For the largest metrically regular set the solution is not so trivial. The following theorem [17] reduces the general problem to a special case:

Theorem 3. Let $A, B \subset \mathbb{F}_2^n$ be a pair of metrically regular sets, i.e. $A = \widehat{B}$, $B = \widehat{A}$. Then there exists a pair of metrically regular sets A^*, B^* at distance 1 from each other such that either $A \subseteq A^*$, $B \subseteq B^*$ or both A,$B \subseteq A^*$.

The theorem tells us that for each metrically regular set in the Boolean cube there exists a metrically regular superset with the covering radius of 1. Therefore, the covering radius of the largest metrically regular set in the Boolean cube is equal to 1. Since for any set A with $\rho(A) = 1$ it holds $A \cup \widehat{A} = \mathbb{F}_2^n$, the largest metrically regular set is the metric (and ordinary) complement of the smallest metrically regular set with the covering radius equal to 1.

The problem is reduced further by the following fact [17]:

Proposition 2. If $C \subseteq \mathbb{F}_2^n$ is a minimal covering code of radius 1, then C is metrically regular.

It follows from the Proposition 2 that any smallest covering code of radius 1 is also a smallest metrically regular set with the covering radius 1. Combined with Theorem 3, this shows that the problem of finding the largest metrically regular set is equivalent to the problem of finding the smallest covering code of radius 1. This is an open problem of coding theory [3] and is solved mostly for particular cases and small dimensions.

Proposition 2 is conjectured [17] to hold true for larger values of the covering radius, however, this has not been proved yet:

Conjecture 1. If $C \subseteq \mathbb{F}_2^n$ is a covering code of radius r of minimal size, then C is metrically regular.

The conjecture was computationally checked [17] for several minimal covering codes with $n = 2r + 3, n = 2r + 4$, where r equals 2 or 3. Constructions of these codes can be found in [2, 5].

3.2. Fixed distances

As we see from the previous subsection, the general problems of finding the largest and the smallest metrically regular sets are reduced to the cases when the covering radius is trivial (equal to either 1 or n). However, the set of bent functions in m variables $\mathcal{B}_m$ has the covering radius equal to $2^{m-1} - 2^{\frac{m}{2}-1}$. In the work [17], the sizes of the sets at a fixed distance r from each other are considered. Theses sizes are estimated nondirectly, through estimating the size of the union of two metrically regular sets, maximally distant one from another. Let us return to the general finite metric space M with a metric $d(\cdot, \cdot)$ admitting values from a set D. Then, the following bound holds [17]:

Theorem 4. Let A,$B \subseteq M$ be a pair of metrically regular sets at distance $r \in D$ from each other, and let C_k be the size of the largest sphere of radius $k \in D$ in M. Then

$$|A| + |B| \geqslant \frac{2|M|}{1 + \sum\limits_{\substack{k \in D \\ k < r}} C_k}.$$

This bound is very similar to the sphere-packing bound on the size of a code, well-known in the coding theory. In the case when the space M is $\mathbb{F}_2^n$ with the Hamming metric, the bound becomes:

Corollary 1. Let $A,B \subseteq \mathbb{F}_2^n$ be a pair of metrically regular sets at distance r from each other. Then

$$|A| + |B| \geqslant \frac{2^{n+1}}{1 + \sum\limits_{k=0}^{r-1} \binom{n}{k}}.$$

4. Strongly metrically regular sets

4.1. Preliminaries

Metrically regular sets are defined by their outstanding metric properties, but a lot of them possess even more regularity. In order to investigate largest and smallest metrically regular sets further, the notion of a *strongly metrically regular* set was introduced in [18].

Let $A \subseteq \mathbb{F}_2^n$ be a set with the covering radius r. The set A is called *strongly metrically regular*, if for any vector $x \in \mathbb{F}_2^n$ it holds

$$d(x, A) + d(x, \widehat{A}) = r.$$

In other words, any vector of the Boolean cube belongs to some shortest path from the set A to the set $\widehat{A}$. It is clear from the definition that any strongly metrically regular set is metrically regular.

The following pair of metrically regular sets gives us a simple example: $A = \{\mathbf{0}\}, \widehat{A} = \{\mathbf{1}\}$. Any vector $x \in \mathbb{F}_2^n$ with the Hamming weight k is at distance k from the set A and at distance $(n-k)$ from the set $\widehat{A}$, so the sum of both distances is equal to n, which is the covering radius of these sets.

But not all metrically regular sets are strongly metrically regular. One of the problems of the international cryptographic olympiad NSUCRYPTO 2016 [26] was to find a metrically regular set which is not strongly metrically regular (or prove that such set does not exist), and several contestants managed to find a solution. The smallest known example of such a set is contained in the Boolean cube of dimension 7.

Let A be an arbitrary subset of the Boolean cube $\mathbb{F}_2^n$. The *layer representation* of $\mathbb{F}_2^n$ with respect to the set A is the sequence of layers defined as follows:

$$A_k := \{x \in \mathbb{F}_2^n | d(x, A) = k\},\ k = 0, 1, \ldots, r$$

where r is the covering radius of A. Using layer representation, strongly metrically regular sets can alternatively be defined as follows [18]:

Proposition 3. Set A is strongly metrically regular if and only if for any k from 0 to r it holds $A_k = \widehat{A}_{r-k}$, where r is the covering radius of both sets.

It is easy to see that completely regular codes [15] are strongly metrically regular. The converse is not true: an example of a strongly metrically regular set which is not a completely regular code is the set $A = \{(000), (011), (111)\}$ in $\mathbb{F}_2^3$.

4.2. Iterative constructions

In the work [18] several iterative constructions of strongly metrically regular sets are obtained:

Theorem 5. Let A be a strongly metrically regular set with the covering radius r. Then $C = A \cup \widehat{A}$ is also a strongly metrically regular set.

This theorem is then generalized to obtain more iterative constructions of strongly metrically regular sets:

Theorem 6. Let A be a strongly metrically regular set with the covering radius $r > 0$ (case $r = 0$ is trivial). Let $i_1, \ldots, i_s$ be a sequence of indices satisfying $0 \leqslant i_1 < i_2 < \ldots < i_{s-1} < i_s \leqslant r$. Then the union $C = \bigcup\limits_{k=1}^{s} A_{i_k}$ is a strongly metrically regular set if and only if there exists a number $\rho > 0$ such that all the following conditions are satisfied:

1) for any $k \in \{1, \ldots, s-1\}$ the distance $(i_{k+1} - i_k)$ is equal to 1, 2ρ or $2\rho + 1$;
2) for any $k \in \{2, \ldots, s-1\}$ at least one of the distances $(i_{k+1} - i_k), (i_k - i_{k-1})$ is greater than 1;
3) i_1 is equal either to ρ or to 0, and if $i_1 = 0$, then $i_2 - i_1 = 2\rho$ or $2\rho + 1$ in case if i_2 exists;
4) i_s is equal either to $r - \rho$ or to r, and if $i_s = r$, then $i_s - i_{s-1} = 2\rho$ or $2\rho + 1$ in case if i_{s-1} exists;

The number ρ is the covering radius of C.

Theorem 6 allows one to construct many new strongly metrically regular sets with smaller covering radii given a strongly metrically regular set with the covering radius r. For example, consider a strongly metrically regular set with the covering radius 20. Then, if we take the union of layers with indices $\{2, 3, 7, 12, 16, 20\}$, it will be a strongly metrically regular set with the covering radius 2 and its metric complement will consist of layers with indices $\{0, 5, 9, 10, 14, 18\}$.

The number of strongly metrically regular sets with the covering radius r which can be constructed using Theorem 6 is also calculated in [18]:

Theorem 7. Let A be a strongly metrically regular set with the covering radius r. Then the number $G_\rho(r)$ of different strongly metrically regular sets with covering radius ρ that can be obtained by applying Theorem 6 to the set A can be calculated using the following recurrent formulas:

$$G_\rho(r) = \begin{cases} G_\rho(r-\rho) + G_\rho(r-\rho-1), & \text{when } r > \rho \\ 2, & \text{when } r = \rho \\ 0, & \text{when } 0 \leqslant r < \rho \end{cases}$$

4.3. Special constructions and lower bounds

Utilizing Theorem 6 and other considerations, two families of "large" strongly metrically regular sets $\{Y_n^r\}$, $\{Z_n^r\}_{n \geqslant 2r}$ are constructed in [18]. Here $Y_n^r, Z_n^r \subseteq \mathbb{F}_2^n$ and $\rho(Y_n^r) = \rho(Z_n^r) = r$. Sets from these families asymptotically cover a large part of the Boolean cube:

$$|Y_n^r| \overset{n\to\infty}{\sim} \frac{2}{2r+1} 2^n. \tag{1}$$

$$|Z_n^r| = 2^{n-2r} \binom{2r}{r} \overset{r\to\infty}{\sim} \frac{1}{\sqrt{\pi r}} 2^n. \tag{2}$$

The lower bound on the sizes of sets from the family $\{Y_n^r\}$ is obtained, which results in the following lower bound on the size of the largest metrically regular set for fixed covering radius:

Theorem 8. Let A be the largest metrically regular set with the covering radius r in the Boolean cube of dimension n ($n \geqslant 2r$), and let ρ be the remainder of $n + 1$ divided by $2r + 1$. Then

$$|A| \geqslant \max\left\{ 2^n \left(\frac{2}{2r+1} - \frac{2}{\sqrt{n-\rho+1}} \right), 2^{n-2r} \binom{2r}{r} \right\} \tag{3}$$

Construction of the family of strongly metrically regular sets $\{Y_n^r\}$ allows one to obtain metrically regular sets with the covering radius r that cover roughly the fraction $\frac{2}{2r+1}$ of the whole Boolean cube when n is big enough, while the family $\{Z_n^r\}$ contains metrically regular sets with the covering radius r that cover roughly the fraction $\frac{1}{\sqrt{\pi r}}$ of the Boolean cube for large values of r.

5. Metric complements and metric regularity of linear codes

5.1. General results

The papers [16, 19] touch upon the topic of metric complements of linear codes in the Boolean cube. First, let us formulate some basic results:

Proposition 4. Let $L \subseteq \mathbb{F}_2^n$ be a linear code. Then the metric complement of L is the union of cosets of L.

This result follows directly from the equality $d_H(x, y) = wt(x + y)$ and the linearity of the code. The following bound is also a simple and well-known result:

Proposition 5. Let $L \subseteq \mathbb{F}_2^n$ be a linear code of dimension k. Then $\rho(L) \leqslant n - k$.

The work [16] describes sufficient and necessary conditions on an arbitrary linear code L to attain this bound, as well as some sufficient conditions for $\rho(L) = n - k - 1$ or $\rho(L) = n - k - 2$. Both of these results also present explicit form of the metric complement of the linear code in question, and in the case when $\rho(L) = n - k$, the code L is found to be metrically regular.

The following characterization of the second metric complement using the first is also presented in [16, 24]:

Proposition 6. Let $L \subseteq \mathbb{F}_2^n$ be a linear code. Then $\rho(\widehat{L}) = \rho(L)$ and a vector x is in $\widehat{\widehat{L}}$ is and only if $x + \widehat{L} = \widehat{L}$.

Corollary 2. Let $L \subseteq \mathbb{F}_2^n$ be a linear code. Assume that $\widehat{L}$ is an affine subspace, i.e. $\widehat{L} = a + L_1$ for some linear code L_1. Then $\widehat{\widehat{L}} = L_1$.

5.2. Sets of affine/bent functions

Let us remember that the notion of a metrically regular set and the problem of investigating and classifying metrically regular sets was first posed by Tokareva in the work [24] when studying metric properties of bent functions, particularly, the duality between bent functions and affine functions.

A Boolean function in even number of variables m is called a *bent function*, if it is at the maximal possible distance from the set of affine functions $\mathcal{A}_m$. If we denote the set of bent functions as $\mathcal{B}_m$, then we have, by definition, $\mathcal{B}_m = \widehat{\mathcal{A}_m}$.

Despite the fact that all characterizations of the set of bent functions that are currently known are rather ineffective when it comes to counting and constructing bent-functions, it turned out that these characterizations are enough to establish metric regularity of the set of affine/bent functions.

It follows from Proposition 6 of the previous subsection that a linear code is metrically regular if and only if no vectors other that those from the code keep its metric complement stable under addition. This property of linear codes was used in [23, 24] to establish that the set of affine functions is the metric complement of the set of bent functions: Tokareva has shown that for any non-affine function f there exists a bent function g (from the so-called "Maiorana-McFarland" class of bent functions) such that $f + g$ is not a bent function. Thus, the following holds:

Theorem 9. Sets of affine functions $\mathcal{A}_m$ and bent functions $\mathcal{B}_m$ are metrically regular.

A. Kutsenko studied metric properties of two subclasses of bent functions called *self-dual* and *anti-self-dual* bent functions. In the work [11] he shows that the set of self-dual bent functions is the metric complement of the set of anti-self-dual bent-functions and vice versa, thus establishing the metric regularity of both of these sets. Other metric properties of bent functions (e.g. the graph of minimal distances between bent functions) were also studied by N. Kolomeec in [7–10].

5.3. Reed-Muller codes

Let $\mathcal{F}^m$ be the set of all Boolean functions in m variables. The Reed-Muller code of order k in m variables is defined as follows:

$$\mathcal{RM}(k,m) = \{f \in \mathcal{F}^m : \deg(f) \leqslant k\},$$

where $\deg(\cdot)$ denotes the degree of the *algebraic normal form (ANF)* [25] of the function. These codes may also be represented as sets of *value vectors* of corresponding functions: binary vectors of length 2^m, containing values which a function assumes on all vectors of $\mathbb{F}_2^m$, listed in some fixed order. Distances between functions can therefore be defined as distances between their value vectors.

The Reed-Muller code of order 1 is, by definition, the set of affine functions, which is, in the case of even number of variables m, metrically regular (as is its metric complement — the set of bent functions). Does this hold for other codes from this family? The work [19] is devoted to the investigation of this metric property for other Reed-Muller codes.

In the work [1], Berlekamp and Welch presented a partition of all cosets of the $\mathcal{RM}(1,5)$ code into 48 classes with respect to the EA-equivalence (extended affine equivalence) and obtained weight distributions for each class of cosets. The full weight distribution allows one to explicitly describe the metric complement of the code. Proposition 6 from the previous subsection is then used to establish the metric regularity of $\mathcal{RM}(1,5)$ in the work [19]. It is shown that for any equivalence class of cosets (other than the $\mathcal{RM}(1,5)$ itself), adding a function from that class to some function from the metric complement $\widehat{\mathcal{RM}}(1,5)$ yields a function outside of the metric complement, leading to the following

Theorem 10. The code $\mathcal{RM}(1,5)$ is metrically regular.

Reed-Muller codes of orders 0, m and $m-1$ coincide with the repetition code, the whole space and the even weight code respectively. It is trivial that all of them are metrically regular. Metric regularity of the Reed-Muller code of order $m-2$ is also easy to establish as follows [19]:

The Reed-Muller code of order $m-2$ has covering radius 2 [3]. By definition, it consists of all Boolean functions of degree at most $m-2$. Since all functions of degree m have odd weights, and all functions of smaller degree have even weights, functions of degree m are at distance 1 from $\mathcal{RM}(m-2,m)$, while functions of degree $m-1$ are at distance 2 and therefore

$$\widehat{\mathcal{RM}}(m-2,m) = \mathcal{RM}(m-1,m) \setminus \mathcal{RM}(m-2,m).$$

Since $\mathcal{RM}(m-2,m)$ is linear, $\rho(\widehat{\mathcal{RM}}(m-2,m)) = \rho(\mathcal{RM}(m-2,m)) = 2$ and thus functions of degree m are at distance 1 from $\widehat{\mathcal{RM}}(m-2,m)$. It follows that $\widehat{\widehat{\mathcal{RM}}}(m-2,m) = \mathcal{RM}(m-2,m)$ and therefore the following holds:

Theorem 11. Codes $\mathcal{RM}(k,m)$ for $k \geqslant m-2$ are metrically regular.

Codes of order $m-3$ are harder to handle. In 1979, McLoughlin [12] has proved that

$$\rho(\mathcal{RM}(m-3,m)) = \begin{cases} m+1, & \text{if } m \text{ is odd,} \\ m+2, & \text{if } m \text{ is even.} \end{cases}$$

This result is reestablished by Cohen et al. in the book "Covering codes" [3] using a method of syndrome matrices, different from the method in [12]. This method allows the author of [19] not only to obtain the covering radius of the Reed-Muller code of order $m-3$, but also to describe the metric complement of this code. As with the covering radius, the cases of even and odd m are distinct.

In the case of even number of variables m, the metric complement can be described as follows:

$$\widehat{\mathcal{RM}}(m-3,m) = \bigcup_{g \in G} (g + \mathcal{RM}(m-3,m)),$$

where

$$G = \{g : \operatorname{supp}(g) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_m, \mathrm{x}_1 + \ldots + \mathrm{x}_m\}, \\ \{\mathrm{x}_1, \ldots, \mathrm{x}_m\} \text{ are linearly independent}\},$$

while for m odd, the description is as follows:

$$\widehat{\mathcal{RM}}(m-3,m) = \bigcup_{g \in G_1 \cup G_2} (g + \mathcal{RM}(m-3,m)),$$

where

$$G_1 = \{g : \operatorname{supp}(g) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_m\}, \{\mathrm{x}_1, \ldots, \mathrm{x}_m\} \text{ are linearly independent}\},$$

and

$$G_2 = \{g : \operatorname{supp}(g) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_{m-1}, \mathrm{x}_1 + \ldots + \mathrm{x}_{m-1}\}, \\ \{\mathrm{x}_1, \ldots, \mathrm{x}_{m-1}\} \text{ are linearly independent}\}.$$

Then, the metric regularity of $\mathcal{RM}(m-3,m)$ is proved by establishing that no functions other that those contained in $\mathcal{RM}(m-3,m)$ preserve the metric complement under addition (once again utilizing Proposition 6 from Subsection 5.1).

The author then considers the code $\mathcal{RM}(2,6)$. Using a proper ordering of the values in the value vectors of functions, this code can be presented in the following manner:

$$\mathcal{RM}(2,6) = \{(\mathbf{u}, \mathbf{u}+\mathbf{v}) : \mathbf{u} \in \mathcal{RM}(2,5), \mathbf{v} \in \mathcal{RM}(1,5)\}.$$

Since both $\mathcal{RM}(2,5)$ and $\mathcal{RM}(1,5)$ were shown to be metrically regular, this constructions proves useful and allows the author to establish the metric regularity of the code $\mathcal{RM}(2,6)$ as well. The proof of this result heavily relies on the fact that $\mathcal{RM}(2,6)$ attains the upper bound on the covering radius provided by the $(\mathbf{u}, \mathbf{u}+\mathbf{v})$ construction, i.e. $\rho(\mathcal{RM}(2,6)) = \rho(\mathcal{RM}(2,5)) + \rho(\mathcal{RM}(1,5))$ [21].

Thus, the metric regularity of the codes $\mathcal{RM}(1,5)$, $\mathcal{RM}(2,6)$ and of the codes $\mathcal{RM}(k,m)$ for $k \geqslant m-3$ has been established. Factoring in the result by Tokareva [23], which proves the metric regularity of $\mathcal{RM}(1,m)$ for even m, this covers all infinite families of Reed-Muller codes with known covering radius. The only other Reed-Muller codes with known covering radius, metric regularity of which has not been yet established, are $\mathcal{RM}(1,7)$ [6, 14] and $\mathcal{RM}(2,7)$ [27]. Given these results, the following conjecture is formulated [19]:

Conjecture 2. All Reed-Muller codes $\mathcal{RM}(k,m)$ are metrically regular.

Conclusion

In this work the main published results concerning metric complements and metric regularity are presented. Metric regularity of partition sets of functions is established. General problem of finding smallest metrically regular sets is found to be trivial, while finding the largest is shown to be as hard as finding the smallest covering code of radius 1. For fixed covering radius, a lower bounds on the sum of sizes of metrically regular sets constituting a pair is obtained. Using the notion of strongly metrically regular set, iterative constructions of metrically regular sets are described and the number of sets which can be obtained using these constructions is calculated. Two families of "large" (relative to the size of $\mathbb{F}_2^n$) metrically regular sets with fixed covering radius are constructed, giving the idea of how big the largest metrically regular sets can be. Characterizations of the first and the second metric complements of linear codes are given. Metric regularity of the Reed-Muller codes $\mathcal{RM}(1,m)$ for m even, $\mathcal{RM}(k,m)$ for $k=0$, $k \geqslant m-3$ and of the codes $\mathcal{RM}(1,5)$, $\mathcal{RM}(2,6)$, is established.

REFERENCES

1. *Berlekamp E., Welch N.* Weight distributions of the cosets of the $(32,6)$ Reed-Muller code. IEEE Transactions on Information Theory, 1972, vol. 18, no. 1, pp. 203–207.
2. *Cohen G., Lobstein A., Sloane N.* Further results on the covering radius of codes. IEEE Transactions on Information Theory, 1986, vol. 32, no. 5, pp. 680–694.
3. *Cohen G., Honkala I., Litsyn S., Lobstein A.* Covering codes. Elsevier, 1997, vol. 54. 541 p.
4. *Cusick T. W., Stănică P.* Cryptographic Boolean functions and applications. Academic Press, 2017, 288 p.
5. *Graham R. L., Sloane N.* On the covering radius of codes. IEEE Transactions on Information Theory, 1985, vol. 31, no. 3, pp. 385–401.
6. *Hou X. D.* Covering Radius of the Reed-Muller code $R(1,7)$ – A Simpler Proof. Journal of Combinatorial Theory, Series A, 1996, vol. 74, no. 2, pp. 337–341.
7. *Kolomeec N. A., Pavlov A. V.* Svoystva bent-funktsiy, nakhodyashchikhsya na minimal'nom rasstoyanii drug ot druga [Properties of bent functions which are at minimal distance from each other]. Prikladnaya diskretnaya matematika, 2009, vol. 6, no. 4, pp. 5–20.
8. *Kolomeets N. A.* Enumeration of the bent functions of least deviation from a quadratic bent function. Journal of Applied and Industrial Mathematics, 2012, vol. 6, no. 3, pp. 306–317.
9. *Kolomeec N. A.* Verkhnyaya otsenka chisla bent-funktsiy na rasstoyanii 2^k ot proizvol'noy bent-funktsii ot 2k peremennykh [Upper bound on the number of bent functions which are at distance 2^k from an arbitrary bent function]. Prikladnaya diskretnaya matematika, 2014, vol. 25, no. 3, pp. 28–39.
10. *Kolomeec N.* The graph of minimal distances of bent functions and its properties. Designs, Codes and Cryptography, 2017, vol. 85, no. 3, pp. 395–410.
11. *Kutsenko A.* Metrical properties of self-dual bent functions. Designs, Codes and Cryptography, 2020, vol. 88, no. 1, pp. 201–222.
12. *McLoughlin A. M.* The Covering Radius of the $(m-3)$-rd Order Reed Muller Codes and a Lower Bound on the $(m-4)$-th Order Reed Muller Codes. SIAM Journal on Applied Mathematics, 1979, vol. 37, no. 2, pp. 419–422.
13. *Mesnager S.* Bent Functions: Fundamentals and Results. Springer International Publishing, 2016, 544 p.
14. *Mykkeltveit J.* The covering radius of the $(128,8)$ Reed-Muller code is 56. IEEE Transactions on Information Theory, 1980, vol. 26, no. 3, pp. 359–362.
15. *Neumaier A.* Completely regular codes. Discrete mathematics, 1992, vol. 106, pp. 353–360.

16. *Oblaukhov A. K.* Metric complements to subspaces in the Boolean cube. Journal of Applied and Industrial Mathematics, 2016, vol. 10, no. 3, pp. 397–403.
17. *Oblaukhov A. K.* Maximal metrically regular sets. Siberian Electronic Mathematical Reports, 2018, vol. 15, pp. 1842–1849.
18. *Oblaukhov A.* A lower bound on the size of the largest metrically regular subset of the Boolean cube. Cryptography and Communications, 2019, vol. 11, no. 4, pp. 777–791.
19. *Oblaukhov A.* `https://arxiv.org/abs/1912.10811` — On metric regularity of Reed-Muller codes, 2019.
20. *Rothaus O. S.* On "bent" functions. Journal of Combinatorial Theory, Series A, 1976, vol. 20, no. 3, pp. 300–305.
21. *Schatz J.* The second order Reed-Muller code of length 64 has covering radius 18. IEEE Transactions on Information Theory, 1981, vol. 17, no. 4, pp. 529–530.
22. *Stănică P., Sasao T., Butler J. T.* Distance duality on some classes of Boolean functions. Journal of Combinatorial Mathematics and Combinatorial Computing, 2018.
23. *Tokareva N. N.* The group of automorphisms of the set of bent functions. Discrete Mathematics and Applications, 2010, vol. 20, no. 5–6, pp. 655–664.
24. *Tokareva N.* Duality between bent functions and affine functions. Discrete Mathematics, 2012, vol. 312, no. 3, pp. 666–670.
25. *Tokareva N.* Bent functions: results and applications to cryptography. Academic Press, 2015. 220 p.
26. *Tokareva N., Gorodilova A., Agievich S., Idrisova V., Kolomeec N., Kutsenko A., Oblaukhov A., Shushuev G.* Mathematical methods in solutions of the problems presented at the Third International Students' Olympiad in Cryptography. Prikladnaya Diskretnaya Matematika, 2018, no. 40, pp. 34–58.
27. *Wang Q.* The covering radius of the Reed-Muller code $RM(2, 7)$ is 40. Discrete Mathematics, 2019, vol. 342, no. 12, Article 111625.

OBLAUKHOV Alexey Konstantinovich — PhD student, Sobolev Institute of Mathematics, Novosibirsk State University, Laboratory of Cryptography JetBrains Research, Novosibirsk, Russia. E-mail: **oblaukhov@gmail.com**

Noname manuscript No.
(will be inserted by the editor)

On metric regularity of Reed-Muller codes

Alexey Oblaukhov

Received: date / Accepted: date

Abstract In this work we study metric properties of the well-known family of binary Reed-Muller codes. Let A be an arbitrary subset of the Boolean cube, and $\widehat{A}$ be the metric complement of A — the set of all vectors of the Boolean cube at the maximal possible distance from A. If the metric complement of $\widehat{A}$ coincides with A, then the set A is called a *metrically regular set*. The problem of investigating metrically regular sets appeared when studying *bent functions*, which have important applications in cryptography and coding theory and are also one of the earliest examples of a metrically regular set. In this work we describe metric complements and establish metric regularity of the codes $\mathcal{RM}(0, m)$ and $\mathcal{RM}(k, m)$ for $k \geqslant m - 3$. Additionally, metric regularity of the $\mathcal{RM}(1, 5)$ code is proved. Combined with previous results by Tokareva N. (2012) concerning duality of affine and bent functions, this proves metric regularity of most Reed-Muller codes with known covering radius. It is conjectured that all Reed-Muller codes are metrically regular.

Keywords metrically regular set · metric complement · covering radius · bent function · Reed-Muller code · deep hole

A. Oblaukhov
Mathematical Center in Akademgorodok,
Sobolev Institute of Mathematics,
Novosibirsk State University,
Laboratory of Cryptography JetBrains Research,
Novosibirsk, Russia
E-mail: oblaukhov@gmail.com
The work was supported by the Russian Foundation for Basic Research (projects no. 18-07-01394, 18-31-00479, 19-31-90093); by the program of basic research of the SB RAS no. I.5.1, project no. 0314-2019-0017; by the Mathematical Center in Akademgorodok, by the Regional Mathematical Center of NSU and by the Laboratory of Cryptography JetBrains Research.

1 Introduction

The problem of investigating and classifying *metrically regular sets* was posed by Tokareva [14,15] when studying metric properties of *bent functions* [11]. A Boolean function f in even number of variables m is called a *bent function* if it is at the maximal possible distance from the set of affine functions. Thus, the set of bent functions $\mathcal{B}_m$ is the metric complement of the set of affine functions $\mathcal{A}_m$, or, in other words, the metric complement of the Reed-Muller code $\mathcal{RM}(1, m)$. It has been proved by Tokareva [14] that the set of affine functions is, conversely, the metric complement of the set of bent functions. It follows that both of these sets are metrically regular, which establishes metric regularity of the codes $\mathcal{RM}(1, m)$ for even m.

It is straightforward from Neumaier's definition [7] of completely regular codes that they are metrically regular (but the converse is not true). Metric regularity of several classes of *partition set functions* is studied in [13], while the work [4] touches upon metric properties of self-dual bent functions. Metric regularity has been actively investigated by the author: metric complements of linear subspaces of the Boolean cube are studied in the paper [8], while the works [9] and [10] are studying possible sizes of the largest and smallest metrically regular set.

In this work we investigate metric properties of Reed-Muller codes. Among the codes of high order, covering radii of the codes $\mathcal{RM}(k, m)$, for $k \geqslant m - 3$ are known. The covering radius of $\mathcal{RM}(1, m)$ for odd $m > 7$ is unknown, but has been determined for $\mathcal{RM}(1, 5)$ [1] and $\mathcal{RM}(1, 7)$ [6,3]. In [12], Schatz has found the covering radius of $\mathcal{RM}(2, 6)$, while recently Wang has established the covering radius of $\mathcal{RM}(2, 7)$ [16]. For $m > 9$, the covering radius of $\mathcal{RM}(2, m)$ is still unknown. We prove that the codes $\mathcal{RM}(k, m)$, for $k = 0$ and $k \geqslant m - 3$ and the code $\mathcal{RM}(1, 5)$ are metrically regular and also describe their metric complements.

The paper is structured as follows. After providing necessary definitions and examples, we prove metric regularity of the $\mathcal{RM}(1, 5)$ code. After that we establish metric regularity of Reed-Muller codes of order 0, order $m - 2$ and higher, and then we move onto the codes of order $m - 3$. In order to handle this case, we describe a "syndrome matrices method" of calculating distances from vectors to the punctured $\mathcal{RM}(m-3, m)$ code, based on the "Covering codes" book by Cohen et al [2]. Following the book, we calculate the covering radius of the Reed-Muller code of order $m-3$. Utilizing the method further, we obtain the metric complement of this code. The description of the complement allows us to establish that only the functions from $\mathcal{RM}(m-3, m)$ are contained in the second metric complement, which proves metric regularity of Reed-Muller codes of order $m - 3$. The paper concludes with the overview of the results obtained and a hypothesis regarding metric regularity of all Reed-Muller codes.

2 Definitions and examples

Let $\mathbb{F}_2^n$ be the space of binary vectors of length n with the Hamming metric. The *Hamming distance* $d(\cdot, \cdot)$ between two binary vectors is defined as the number of coordinates in which these vectors differ, while $wt(\cdot)$ denotes the *weight* of a vector, i.e. the number of nonzero values it contains. The plus sign + will denote addition

modulo two (componentwise in case of vectors), while the componentwise product of two binary vectors will be denoted as $*$.

Let $X \subseteq \mathbb{F}_2^n$ be an arbitrary set and $y \in \mathbb{F}_2^n$ be an arbitrary vector. The distance from the vector y to the set X is defined as

$$d(y, X) = \min_{x \in X} d(y, x).$$

The *covering radius* of the set X is defined as

$$\rho(X) = \max_{z \in \mathbb{F}_2^n} d(z, X).$$

The set X with $\rho(X) = r$ is also called a *covering code* [2] of radius r.

Consider the set

$$Y = \{y \in \mathbb{F}_2^n | d(y, X) = \rho(X)\}$$

of all vectors at the maximal possible distance from the set X. This set is called the *metric complement* [8] of X and is denoted by $\widehat{X}$. Vectors from the metric complement are sometimes called *deep holes* of a code. If $\widehat{\widehat{X}} = X$ then the set X is said to be *metrically regular* [15].

Note that metrically regular sets always come in pairs, i.e. if A is a metrically regular set, then its metric complement $\widehat{A}$ is also a metrically regular set and both of them have the same covering radius. For some simple examples of metric complements and metrically regular sets, refer to [8–10].

Let $\mathcal{F}^m$ be the set of all Boolean functions in m variables. Reed-Muller code of order k is defined as:

$$\mathcal{RM}(k, m) = \{f \in \mathcal{F}^m : \deg(f) \leqslant k\},$$

where $\deg(\cdot)$ denotes the degree of the *algebraic normal form (ANF)* of the function. These codes may be also represented as sets of *value vectors* of functions. Throughout the paper we will often switch between these two representations. In most cases, m will denote the number of variables, while $n := 2^m$ will denote the dimension of the space of value vectors, which have coordinates numbered from 0 to $2^m - 1$. The i-th coordinate of a value vector is the value of the corresponding function at the binary vector of length m which is a binary representation of the number i. Weights of functions, distances between functions and between a function and a set of functions are defined as distances between their value vectors.

From now on, vectors of length m and square $m \times m$ matrices will be denoted using roman typestyle letters (e.g. x, A), while vectors of length n and vectors derived from them, as well as matrices related to such vectors, will be denoted using bold letters (e.g. $\mathbf{v}, \mathbf{B}$).

Let f and g be two functions in m variables. Denote as $L_{\mathrm{A}}^{\mathrm{b}} : \mathbb{F}_2^m \to \mathbb{F}_2^m$ the affine transformation of the variables with the matrix A and the vector b):

$$(f \circ L_{\mathrm{A}}^{\mathrm{b}})(\mathrm{x}) = f(\mathrm{Ax} + \mathrm{b}).$$

Here $\circ$ denotes the composition of the functions. If the vector b is zero, it will be omitted from the notation. Functions f and g are called *linearly equivalent* if one can be obtained from the other by applying a nonsingular linear transformation to the variables, i.e. $f = g \circ L_{\mathrm{A}}$, where $\det \mathrm{A} \neq 0$.

Extended affine equivalence is more common when classifying boolean functions: functions f and g are called EA-*equivalent* if there exists a nonsingular linear transformation of variables A, a boolean vector b of length m and a function l of degree at most 1 such that $f = g \circ L_{\mathrm{A}}^{\mathrm{b}} + l$.

For our study we will use a variant of these two equivalence relations, which will be referred to as *extended linear equivalence (to the power of k)*. Functions f and g are called EL^k-*equivalent* if there exists a nonsingular binary matrix A and a function c of degree at most k such that

$$f = g \circ L_{\mathrm{A}} + c.$$

It is easy to see that this relation is indeed an equivalence. We will denote this equivalence as $f \overset{k}{\sim} g$.

Reed-Muller code of order k in m variables is usually denoted as $\mathcal{RM}(k, m)$. Since we will refer to these codes regularly, we will instead use $\mathcal{R}_k$ to denote a Reed-Muller code of order k in m variables.

3 Reed-Muller code $\mathcal{RM}(1, 5)$

In the work [1], Berlekamp and Welch presented a partition of all cosets of the $\mathcal{RM}(1, 5)$ code into 48 classes with respect to the EA-equivalence and obtained weight distributions for each class of cosets. Four of these cosets contain only codewords of weight 12 and higher, and those cosets constitute the metric complement of $\mathcal{RM}(1, 5)$. Thus we can present the metric complement of this code as:

$$\widehat{\mathcal{RM}}(1, 5) = \{f : f \overset{\mathrm{EA}}{\sim} g \text{ for some } g \text{ from one of 4 farthest classes}\}$$

Since $\mathcal{RM}(1, 5)$ is linear, it follows [8] that

$$\rho(\widehat{\mathcal{RM}}(1, 5)) = \rho(\mathcal{RM}(1, 5)) = 12$$

and $f \in \widehat{\widehat{\mathcal{RM}}}(1, 5)$ if and only if $f + \widehat{\mathcal{RM}}(1, 5) = \widehat{\mathcal{RM}}(1, 5)$. Thus, in order to establish metric regularity of $\mathcal{RM}(1, 5)$, we must prove that for every $f \notin \mathcal{RM}(1, 5)$ it holds $f + \widehat{\mathcal{RM}}(1, 5) \neq \widehat{\mathcal{RM}}(1, 5)$.

Let $f_c \notin \mathcal{RM}(1, 5)$ be a function from a certain coset equivalence class C. Assume that the function $f_c + g_c$, where $g_c \in \widehat{\mathcal{RM}}(1, 5)$, does not belong to any of the 4 equivalence classes from the complement $\widehat{\mathcal{RM}}(1, 5)$. This implies that $f_c + \widehat{\mathcal{RM}}(1, 5) \neq \widehat{\mathcal{RM}}(1, 5)$ and thus f_c is not in the second metric complement.

Let now $f \notin \mathcal{RM}(1, 5)$ be an arbitrary function from the class C, and let $(\mathrm{A}, \mathrm{b}, l)$ be the matrix, the vector and the affine function such that

$$f \circ L_{\mathrm{A}}^{\mathrm{b}} + l = f_c.$$

Denote

$$g_f = (g_c + l) \circ L_{\mathrm{A}^{-1}}^{\mathrm{A}^{-1}\mathrm{b}}.$$

Then the function $f + g_f$ is EA-equivalent to $f_c + g_c$ and therefore does not belong to $\widehat{\mathcal{RM}}(1, 5)$. Since $g_f \in \widehat{\mathcal{RM}}(1, 5)$, this implies that $f \notin \widehat{\widehat{\mathcal{RM}}}(1, 5)$.

Thus, if we prove that $f + g \notin \widehat{\mathcal{RM}}(1,5)$ for some $f \in C$ and some $g \in \widehat{\mathcal{RM}}(1,5)$, we will prove that no function from the equivalence class C is in the second metric complement.

The list of all representatives of equivalence classes of $\mathcal{RM}(1,5)$ and the proof that none of the classes, except for $\mathcal{RM}(1,5)$ itself, belong to the second metric complement can be found in the Appendix I of the paper under the

Theorem 1 *The code $\mathcal{RM}(1,5)$ is metrically regular.*

4 Reed-Muller codes of orders 0, m, $m-1$ and $m-2$

Reed-Muller codes of orders 0, m and $m-1$ coincide with the repetition code, the whole space and the even weight code respectively. It is trivial that all of them are metrically regular.

The Reed-Muller code of order $m-2$ has covering radius 2 [2]. By definition, it consists of all Boolean functions of degree at most $m-2$. Since all functions of degree m have odd weight, and all functions of smaller degree have even weight, functions of degree m are at distance 1 from $\mathcal{R}_{m-2}$, while functions of degree $m-1$ are at distance 2 and therefore

$$\widehat{\mathcal{R}}_{m-2} = \mathcal{R}_{m-1} \setminus \mathcal{R}_{m-2}.$$

Since $\mathcal{R}_{m-2}$ is linear, $\rho(\widehat{\mathcal{R}}_{m-2}) = \rho(\mathcal{R}_{m-2}) = 2$ and thus functions of degree m are at distance 1 from $\widehat{\mathcal{R}}_{m-2}$. It follows that $\widehat{\widehat{\mathcal{R}}}_{m-2} = \mathcal{R}_{m-2}$ and $\mathcal{R}_{m-2}$ is metrically regular.

5 Reed-Muller codes of order $m-3$: Syndrome method

McLoughlin [5] has proved that

$$\rho(\mathcal{R}_{m-3}) = \begin{cases} m+1, & \text{if } m \text{ is odd,} \\ m+2, & \text{if } m \text{ is even.} \end{cases}$$

We are going to reestablish this result following the "Covering codes" book by Cohen et al, since our new results that follow rely on methods and terminology described in the book. In particular, we will describe the method of obtaining the covering radius of $\mathcal{R}_{m-3}$ using syndrome matrices as it is presented in the book, with few minor adjustments. After that we will proceed to study the metric complement of $\mathcal{R}_{m-3}$. Results in Chapters 5 and 6, as well as general result concerning the covering radius of $\mathcal{R}_{m-3}$, belong to Cohen et al [2], while all subsequent results concerning metric complements and metric regularity of the codes have been obtained by the author.

Let us first consider the covering radius of the punctured Reed-Muller code $\mathcal{R}^{\circ}_{m-3}$, i.e., the code without the 0-th coordinate (which corresponds to the value of the function at zero). Let us denote the parity check matrix of this code as $\mathbf{H}$. The matrix $\mathbf{H}$ coincides with the parity check matrix of the non-punctured code

$\mathcal{R}_{m-3}$, but with the first all-one row and the first column removed. Since $\mathcal{R}_{m-3}$ is dual to the code $\mathcal{R}_2$, the rows of $\mathbf{H}$ are punctured value vectors of the functions

$$x_1, \ldots, x_m, x_1x_2, x_1x_3, \ldots, x_{m-1}x_m.$$

The syndrome $\mathbf{s}$ of an arbitrary vector $\mathbf{v} \in \mathbb{F}_2^{n-1}$ is the product $\mathbf{Hv^T}$. Let us present the syndrome $\mathbf{s}$ as an $m \times m$ symmetric matrix $\mathbf{S}$, where element $s_{i,j}$ of the matrix is equal to the component of the syndrome corresponding to the row x_ix_j of the parity check matrix $\mathbf{H}$. Diagonal element $s_{i,i}$ of this matrix corresponds to the row x_i of the matrix $\mathbf{H}$. Thus we have built a one-to-one correspondence between cosets of $\mathcal{R}^\circ_{m-3}$ and syndrome matrices $\mathbf{S}$, i.e., all symmetric binary matrices.

Let $\mathbf{e_1^\circ}, \ldots \mathbf{e_m^\circ} \in \mathbb{F}_2^{n-1}$ be the punctured value vectors of the functions $x_1, \ldots, x_m$. Notice that the row of $\mathbf{H}$ corresponding to the function x_ix_j is the componentwise product $\mathbf{e_i^\circ} * \mathbf{e_j^\circ}$.

Consider an $m \times (n-1)$ matrix $\mathbf{B_v}$ which has $\mathbf{e_i^\circ} * \mathbf{v}$ as its i-th row. Then the symmetric matrix $\mathbf{S_v} = \mathbf{B_v B_v^T}$ corresponds to the syndrome $\mathbf{Hv^T}$ of the vector $\mathbf{v}$. It is easy to see that if $f_\mathbf{v}$ is a function with the punctured value vector $\mathbf{v}$, then the set of nonzero columns of $\mathbf{B_v}$ is precisely the support of $f_\mathbf{v}$ (bar, possibly, the zero vector). The number of nonzero columns in $\mathbf{B_v}$ is equal to the weight of the vector $\mathbf{v}$.

Given an arbitrary vector $\mathbf{v} \in \mathbb{F}_2^{n-1}$, its distance $d(\mathbf{v}, \mathcal{R}^\circ_{m-3})$ from the code is equal to the weight of the coset leader:

$$d(\mathbf{v}, \mathcal{R}^\circ_{m-3}) = \min_{\mathbf{u}:\mathbf{Hu^T}=\mathbf{Hv^T}} wt(\mathbf{u}).$$

Using the established correspondences between syndromes and symmetric matrices, we can rewrite this as

$$d(\mathbf{v}, \mathcal{R}^\circ_{m-3}) = \min_{\mathbf{u}:\mathbf{B_u B_u^T}=\mathbf{S_v}} Col(\mathbf{B_u}),$$

where $Col(\mathbf{B_u})$ is the number of nonzero columns in the matrix $\mathbf{B_u}$. Let us denote this minimum as $t(\mathbf{S}) := \min\limits_{\mathbf{u}:\mathbf{B_u B_u^T}=\mathbf{S}} Col(\mathbf{B_u})$. Then

$$d(\mathbf{v}, \mathcal{R}^\circ_{m-3}) = t(\mathbf{S_v}),$$

and, since the correspondence between all syndromes and all symmetric matrices is one-to-one, we have

$$\rho(\mathcal{R}^\circ_{m-3}) = \max_\mathbf{v} d(\mathbf{v}, \mathcal{R}^\circ_{m-3}) = \max_\mathbf{S} t(\mathbf{S}).$$

Moreover, a vector $\mathbf{v}$ is in the metric complement $\widehat{\mathcal{R}}^\circ_{m-3}$ if and only if $t(\mathbf{S_v}) = \rho(\mathcal{R}^\circ_{m-3})$.

We will call any matrix $\mathbf{B}$ such that $\mathbf{BB^T} = \mathbf{S}$ a *factor* of $\mathbf{S}$. We can thus describe the value $t(\mathbf{S})$ as *the minimum number of nonzero columns in a factor over all factors of* $\mathbf{S}$ *of the form* $\mathbf{B_u}$*, where* $\mathbf{u} \in \mathbb{F}_2^{n-1}$. We will call any factor achieving this minimum a *minimal factor*.

Let us now expand the definition of the value $t(\mathbf{S})$.

Lemma 1 *Let* $\mathbf{S}$ *be a symmetric matrix, and let* $\mathbf{B}$ *be its factor (i.e.* $\mathbf{BB^T} = \mathbf{S}$*). The following operations do not change the property of* $\mathbf{B}$ *being a factor of* $\mathbf{S}$*:*

1. *deleting a zero column;*
2. *deleting two equal columns;*
3. *swapping any two columns;*
4. *adding an arbitrary vector b to each column from some subset of columns of* $\mathbf{B}$ *of even size, given that all columns of this subset sum to zero.*

Proof. Routine, left to the reader. □

Since subsets of nonzero columns of matrices $\{\mathbf{B_u}|\mathbf{u} \in \mathbb{F}_2^{n-1}\}$ are exactly all possible subsets of nonzero columns of length m, and using Lemma 1, we can remove all zero columns from allowed factors and ignore the possibility of duplicate columns and thus reformulate the definition of the value $t(\mathbf{S})$ in the following manner, allowing arbitrary size matrices:

$t(\mathbf{S})$ *is the minimum number of columns in a factor over all factors of* $\mathbf{S}$*. Any factor achieving this minimum is called a minimal factor of* $\mathbf{S}$*.*

Moreover, any factor $\mathbf{B}$ of $\mathbf{S}$ corresponds to exactly one factor of the initial form $\mathbf{B_u}$ — the factor with the set of nonzero columns coinciding with the set of nonzero columns of $\mathbf{B}$. Therefore, presenting a minimal factor for a symmetric matrix $\mathbf{S}$ allows us to obtain a coset leader $\mathbf{u}$ for the coset which this symmetric matrix represents.

6 Reed-Muller codes of order $m - 3$: Covering radius

In order to determine the covering radius of $\mathcal{R}^{\circ}_{m-3}$, let us now investigate the maximum value of $t(\mathbf{S})$. Obviously,

$$t(\mathbf{S}) \geqslant \min_{\mathbf{B}:\mathbf{BB^T}=\mathbf{S}} \operatorname{rank}(\mathbf{B}) \geqslant \operatorname{rank}(\mathbf{S})$$

for any matrix $\mathbf{S}$, and therefore

$$\max_{\mathbf{S}} t(\mathbf{S}) \geqslant m.$$

Notice that, if $\mathbf{S} = \mathbf{BB^T}$, then the vector consisting of all diagonal entries of the matrix $\mathbf{S}$ is the sum of all columns of $\mathbf{B}$. Assume that the matrix $\mathbf{S}$ is nonsingular and has an all-zero diagonal. Then all columns of any of its factor $\mathbf{B}$ sum to zero and thus all nonzero columns form a linearly dependent set of vectors. Since $\operatorname{rank}(\mathbf{B}) \geqslant \operatorname{rank}(\mathbf{S}) = m$, this leads to $t(\mathbf{S}) \geqslant m + 1$. Note that a matrix $\mathbf{S}$ with such properties exists if and only if m is even (see e.g. [2] p. 249).

Combining these bounds, we obtain

$$\max_{\mathbf{S}} t(\mathbf{S}) \geqslant m + 1 - \pi(m),$$

where $\pi(m)$ is the parity function, equal to 1 for odd m and to 0 for even m.

We will now prove that this bound is tight. The following lemma will help us to characterize minimal factors:

Lemma 2 *Let* $\mathbf{S}$ *be a symmetric matrix, and let* $\mathbf{B}$ *be its minimal factor. Then all proper subsets of columns of* $\mathbf{B}$ *are linearly independent.*

Proof. See Appendix II. □

Assume that for some symmetric matrix $\mathbf{S}$ it holds $t(\mathbf{S}) \geqslant m+2$. This means that any minimal factor $\mathbf{B}$ of $\mathbf{S}$ has at least $m+2$ columns and therefore contains a linearly dependent proper subset of columns, which contradicts Lemma 2. Therefore, $t(\mathbf{S}) \leqslant m+1$ for any matrix $\mathbf{S}$.

7 Reed-Muller codes of order $m-3$: Case m is even

7.1 Covering radius and metric complement of the punctured code

Let the number of variables m be even. Combining the upper and the lower bound obtained in the previous chapter, we get:

$$\rho(\mathcal{R}^{\circ}_{m-3}) = \max_{\mathbf{S}} t(\mathbf{S}) = m+1$$

and a vector $\mathbf{v}$ is in the metric complement of $\mathcal{R}^{\circ}_{m-3}$ if and only if $t(\mathbf{S_v}) = m+1$. The following lemma will help us to characterize the metric complement of $\mathcal{R}^{\circ}_{m-3}$:

Lemma 3 *Let $\mathbf{S}$ be a symmetric $m \times m$ matrix, where m is even. Then $t(\mathbf{S}) = m+1$ if and only if $\mathbf{S} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$ for some $m \times (m+1)$ matrix $\mathbf{B}$ of rank m such that all its columns sum to zero.*

Proof. See Appendix II. □

Clearly, this lemma describes all minimal factors of all matrices $\mathbf{S}$ satisfying $t(\mathbf{S}) = m+1$. Let

$$\mathrm{U} = \{\mathbf{u} : \mathbf{B_u} \text{ has } m+1 \text{ nonzero columns, } m \text{ of which are linearly independent and all of them sum to zero}\}.$$

It is easy to see that the set of matrices $\{\mathbf{B_u} | \mathbf{u} \in \mathrm{U}\}$ (up to columns permutations and zero columns removal) includes exactly all minimal factors described in the Lemma 3. Thus, if $t(\mathbf{S}) = m+1$ for some matrix $\mathbf{S}$, then there exists a vector $\mathbf{u} \in \mathrm{U}$ such that $\mathbf{S} = \mathbf{B_u}\mathbf{B_u^T}$. Conversely, for any $\mathbf{u} \in \mathrm{U}$ it holds $t(\mathbf{B_u}\mathbf{B_u^T}) = m+1$. Therefore, vectors from the set U cover all cosets contained in the metric complement $\widehat{\mathcal{R}}^{\circ}_{m-3}$:

$$\widehat{\mathcal{R}}^{\circ}_{m-3} = \bigcup_{\mathbf{u} \in \mathrm{U}} \left(\mathbf{u} + \mathcal{R}^{\circ}_{m-3}\right).$$

7.2 Covering radius and metric complement of the non-punctured code

We have obtained the covering radius and described the metric complement of the punctured code. Let us return to the regular, non-punctured Reed-Muller code $\mathcal{R}_{m-3}$. Since it is obtained from the punctured code by adding a parity check bit at 0-th coordinate, the following result will be of use:

Lemma 4 *Let C be a code with the covering radius r and the metric complement $\widehat{C}$. Let C_π be the code obtained from C by adding a parity check bit. Then $\rho(C_\pi) = r+1$ and $\widehat{C}_\pi$ is obtained from $\widehat{C}$ by*

1. *adding a parity check bit to all vectors in case if r is odd or*
2. *adding a reverse parity check bit to all vectors in case if r is even.*

Proof. See Appendix III. □

Using this lemma we can conclude that the covering radius of the non-punctured Reed-Muller code $\mathcal{R}_{m-3}$ is equal to $m+2$ and its metric complement can be described as follows:

$$\widehat{\mathcal{R}}_{m-3} = \bigcup_{\mathbf{u}\in \mathrm{U}} \left((\pi(\mathbf{u}), \mathbf{u}) + \mathcal{R}_{m-3}\right).$$

Recall that, if $f_{\mathbf{v}}$ is the function with the value vector $\mathbf{v}$ (non-punctured), then the set of nonzero columns of the matrix $\mathbf{B}_{\mathbf{v}^\circ}$ coincides with the support of the function $f_{\mathbf{v}}$, bar, possibly, the zero vector. Considering also that all vectors in U have odd weight and added parity check bit corresponds to the value of the function at the all-zero vector, we can rewrite the metric complement of $\mathcal{R}_{m-3}$ in terms of functions instead of their value vectors:

$$\widehat{\mathcal{R}}_{m-3} = \bigcup_{f\in F} \left(f + \mathcal{R}_{m-3}\right),$$

where

$$F = \{f_{(\pi(\mathbf{u}),\mathbf{u})} : \mathbf{u} \in U\} = \{f : \mathrm{supp}(f) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_m, \mathrm{x}_1 + \ldots + \mathrm{x}_m\}, \\ \{\mathrm{x}_1, \ldots, \mathrm{x}_m\} \text{ are linearly independent}\}.$$

It is easy to see that all functions in F form an equivalence class with respect to linear equivalence. Let us pick any function f^* from this class. We can now say that a function g is in $\widehat{\mathcal{R}}_{m-3}$ if and only if $g = f^* \circ L_{\mathrm{A}} + h$ for some nonsingular matrix A and some function h of degree at most $m-3$.

Recall that functions f and g are EL^k*-equivalent* if there exists a nonsingular binary matrix A and a function h of degree at most k such that $g = f \circ L_{\mathrm{A}} + h$. Therefore, $g \in \widehat{\mathcal{R}}_{m-3}$ if and only if $g \overset{m-3}{\sim} f^*$, where f^* is an arbitrarily chosen representative of the class F. In fact, since all functions in the metric complement are equivalent, we can pick any function from $\widehat{\mathcal{R}}_{m-3}$ as our reference for equivalence (and we will actively change this reference whenever convenient). We will call EL^{m-3}–equivalence just "equivalence" for brevity from now on.

Let us give an explicit (algebraic normal form) description of a certain function from F. Denote as f^* the function with the support $\{0, \mathrm{e}_1, \mathrm{e}_2, \ldots, \mathrm{e}_m, 1\}$, where e_i is the vector with 1 only in the i-th coordinate. Clearly, $f^* \in F$ and it is easy to construct the algebraic normal form of this function: it is the sum of all monomials containing even number of variables, excluding the monomial with all variables included:

$$f^*(\mathrm{x}) = 1 + \sum_{k=1}^{\frac{m}{2}-1} \sum_{1\leqslant i_1<\ldots<i_{2k}\leqslant m} x_{i_1} x_{i_2} \ldots x_{i_{2k}}.$$

This function is equivalent to the sum of all monomials containing $m-2$ variables, so let us use this last function as f^* moving forward. Let $\overline{x_i}$ denote the product of all m variables except x_i, and let $\overline{x_i x_j}$ denote the product of all m variables

except x_i and x_j. Using these conventions, we can write this new representative function as follows:

$$f^*(\mathrm{x}) = \sum_{1 \leqslant i < j \leqslant m} \overline{x_i x_j}.$$

7.3 **Metric regularity**

We have established that

$$\widehat{\mathcal{R}}_{m-3} = \{g : g \overset{m-3}{\sim} f^*\},$$

where f^* is some function from the class F (or from $\widehat{\mathcal{R}}_{m-3}$), and have presented a representative.

Since the code $\mathcal{R}_{m-3}$ is linear, it follows [8] that $\rho(\widehat{\mathcal{R}}_{m-3}) = \rho(\mathcal{R}_{m-3}) = m+2$ and a function f is in $\widehat{\widehat{\mathcal{R}}}_{m-3}$ if and only if $f + \widehat{\mathcal{R}}_{m-3} = \widehat{\mathcal{R}}_{m-3}$. With this in mind, let us prove the metric regularity of $\mathcal{R}_{m-3}$ by proving that no functions other that those contained in $\mathcal{R}_{m-3}$ preserve the metric complement under addition.

Case 1. Let $f \notin \mathcal{R}_{m-3}$ be a function of degree greater than $m-2$. Since EL^{m-3}-equivalence preserves degree of functions with degree higher than $m-3$, any $g \in \widehat{\mathcal{R}}_{m-3}$ has degree $m-2$, while $f+g$ has higher degree and therefore cannot be equivalent to any of the functions from $\widehat{\mathcal{R}}_{m-3}$. Thus, functions of degree greater than $m-2$ do not preserve any function from the metric complement and therefore cannot be in $\widehat{\widehat{\mathcal{R}}}_{m-3}$.

Case 2. Let $f \notin \mathcal{R}_{m-3}$ be a function of degree $m-2$. Let us write it as follows:

$$f(\mathrm{x}) = \sum_{(i,j) \in I} \overline{x_i x_j} + h(\mathrm{x}),$$

where $\deg(h) < m-2$. Denote as $\tilde{f}$ the quadratic function defined by:

$$\tilde{f}(\mathrm{x}) = \sum_{(i,j) \in I} x_i x_j.$$

We will call $\tilde{f}$ the *quadratic dual* of f.

The following result would be of use when handling this case:

Lemma 5 *Let f and g be two function of degree $m-2$. Then $f \overset{m-3}{\sim} g$ if and only if their quadratic duals are* EL^1*-equivalent (*EA*-equivalent).*

Proof. See Appendix III. □

It is known that any quadratic boolean function is EA-equivalent to the function of the form $x_1x_2 + x_3x_4 + \ldots + x_{2k-1}x_{2k}$ for some $k \leqslant \frac{m}{2}$, and any two functions of this form with different number of variables are not EA-equivalent one to the other. Using this result and Lemma 5 we can say that the function f is equivalent to the function p_k for some k ($0 < k \leqslant \frac{m}{2}$), where

$$p_k(\mathrm{x}) = \overline{x_1x_2} + \overline{x_3x_4} + \ldots + \overline{x_{2k-1}x_{2k}} = \sum_{i=1}^{k} \overline{x_{2i-1}x_{2i}}.$$

Let A be the matrix and h be the function of degree at most $m-3$ such that $f \circ L_{\mathrm{A}} + h = p_k$.

Trivially, f^* is equivalent to $p_{\frac{m}{2}}$. Then $f \circ L_{\mathrm{A}} + h + p_{\frac{m}{2}}$ is equivalent to $p_{\frac{m}{2}-k}$, which is (by Lemma 5) not equivalent to $p_{\frac{m}{2}}$ and therefore not equivalent to f^*.

Thus, for an arbitrarily chosen function f of degree $m-2$ we have found a function $g = (h + p_{\frac{m}{2}}) \circ L_{\mathrm{A}^{-1}}$ from the metric complement $\widehat{\mathcal{R}}_{m-3}$ such that $f+g$ is not equivalent to f^*. This means that $f \notin \widehat{\widehat{\mathcal{R}}}_{m-3}$.

Since all functions which are not in $\mathcal{R}_{m-3}$ have degree $m-2$ or higher, we have proven that none of them are in the second metric complement, and therefore $\mathcal{R}_{m-3}$ is metrically regular.

8 Reed-Muller codes of order $m-3$: Case m is odd

8.1 Covering radius and metric complement of the punctured code

Let the number of variables m be odd. Many arguments for this case are similar or identical to the ones for the previous case. The following lemma will be of use:

Lemma 6 *Let $\mathbf{S}$ be a symmetric $m \times m$ matrix, where m is odd. Then $t(\mathbf{S}) \leqslant m$, and the equality is achieved if and only if $\mathbf{S} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$ for some $m \times m$ matrix $\mathbf{B}$ which is either nonsingular, or has rank $m-1$ and all columns summing to zero.*

Proof. See Appendix II. □

From Lemma 6 we can conclude that, in the case of odd m,

$$\rho(\mathcal{R}^{\circ}_{m-3}) = \max_{\mathbf{S}} t(\mathbf{S}) = m,$$

and a vector $\mathbf{v}$ is in the metric complement of $\mathcal{R}^{\circ}_{m-3}$ if and only if $t(\mathbf{S}_{\mathbf{v}}) = m$.

Lemma 6 also describes all minimal factors of all matrices $\mathbf{S}$ satisfying $t(\mathbf{S}) = m$. Let

$$\mathrm{U}_1 = \{\mathbf{u} : \mathbf{B}_{\mathbf{u}} \text{ has } m \text{ nonzero columns which are linearly independent}\}$$

and

$$\mathrm{U}_2 = \{\mathbf{u} : \mathbf{B}_{\mathbf{u}} \text{ has } m \text{ nonzero columns, } m-1 \text{ of which are linearly independent and the sum of all columns is equal to zero}\}.$$

It is easy to see that the set of matrices $\{\mathbf{B}_{\mathbf{u}} | \mathbf{u} \in \mathrm{U}_1 \cup \mathrm{U}_2\}$ (up to columns permutations and zero columns removal) includes exactly all minimal factors described in the Lemma 6. Thus, if $t(\mathbf{S}) = m$ for some matrix $\mathbf{S}$, then there exists a vector $\mathbf{u} \in \mathrm{U}$ such that $\mathbf{S} = \mathbf{B}_{\mathbf{u}}\mathbf{B}_{\mathbf{u}}^{\mathbf{T}}$. Conversely, for any $\mathbf{u} \in \mathrm{U}$ it holds $t(\mathbf{B}_{\mathbf{u}}\mathbf{B}_{\mathbf{u}}^{\mathbf{T}}) = m$. Therefore, vectors from the set U cover all cosets contained in the metric complement $\widehat{\mathcal{R}}^{\circ}_{m-3}$:

$$\widehat{\mathcal{R}}^{\circ}_{m-3} = \bigcup_{\mathbf{u} \in \mathrm{U}_1 \cup \mathrm{U}_2} \mathbf{u} + \mathcal{R}^{\circ}_{m-3}.$$

8.2 **Covering radius and metric complement of the non-punctured code**

We have obtained the covering radius and described the metric complement of the punctured code. Let us return to the regular, non-punctured Reed-Muller code $\mathcal{R}_{m-3}$. Since it is obtained from punctured code by adding a parity check bit, using Lemma 4 we can conclude that the covering radius of $\mathcal{R}_{m-3}$ is equal to $m+1$, and its metric complement is

$$\widehat{\mathcal{R}}_{m-3} = \bigcup_{\mathbf{u}\in \mathrm{U}_1\cup\mathrm{U}_2} (\pi(\mathbf{u}), \mathbf{u}) + \mathcal{R}_{m-3}.$$

Recall once again that, if $f_{\mathbf{v}}$ is the function with a value vector $\mathbf{v}$ (non-punctured), then the set of nonzero columns of $\mathbf{B}_{\mathbf{v}^\circ}$ coincides with the support of the function $f_{\mathbf{v}}$, bar, possibly, the zero vector. Considering also that all vectors in $\mathrm{U}_1 \cup \mathrm{U}_2$ have odd weight and added parity check bit corresponds to the value of the function at the all-zero vector, we can rewrite the metric complement of $\mathcal{R}_{m-3}$ in terms of functions instead of their value vectors:

$$\widehat{\mathcal{R}}_{m-3} = \bigcup_{f\in F_1\cup F_2} f + \mathcal{R}_{m-3},$$

where

$$F_1 = \{f_{(\pi(\mathbf{u}),\mathbf{u})} : \mathbf{u} \in \mathrm{U}_1\} = \\ = \{f : \mathrm{supp}(f) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_m\}, \{\mathrm{x}_1, \ldots, \mathrm{x}_m\} \text{ are linearly independent}\},$$

and

$$F_2 = \{f_{(\pi(\mathbf{u}),\mathbf{u})} : \mathbf{u} \in \mathrm{U}_2\} = \\ = \{f : \mathrm{supp}(f) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_{m-1}, \mathrm{x}_1 + \ldots + \mathrm{x}_{m-1}\}, \\ \{\mathrm{x}_1, \ldots, \mathrm{x}_{m-1}\} \text{ are linearly independent}\}.$$

It is easy to see that all functions in F_1 form an equivalence class with respect to linear equivalence, so do functions in F_2. Let us pick any two functions f_1^*, f_2^* from these two classes, one from each class. We can now say that a function g is in $\widehat{\mathcal{R}}_{m-3}$ if and only if $g = f_1^* \circ L_\mathrm{A} + h$ or $g = f_2^* \circ L_\mathrm{A} + h$ for some nonsingular matrix A and some function h of degree not greater than $m-3$.

Therefore, $g \in \widehat{\mathcal{R}}_{m-3}$ if and only if $g \overset{m-3}{\sim} f_1^*$ or $g \overset{m-3}{\sim} f_2^*$, where f_1^* is an arbitrarily chosen representative of the class F_1 and f_2^* is an arbitrarily chosen representative of the class F_2. In fact, we can pick any function from EL^{m-3}–equivalence class of F_1 and from EL^{m-3}–equivalence class of F_2 respectively as our references of equivalence f_1^* and f_2^*.

Let us give an explicit (algebraic normal form) description of a certain function from F_1. Denote as f_1^* the function with the support $\{0, \mathrm{e}_1, \mathrm{e}_2, \ldots, \mathrm{e}_{m-1}, 1\}$. After a bit of calculation one can explicitly describe its ANF:

$$f_1^*(\mathrm{x}) = \overline{x_m} + (1 + x_m)\left(1 + \sum_{k=1}^{\frac{m-3}{2}} \sum_{1\leqslant i_1<\ldots<i_{2k}\leqslant m-1} x_{i_1}x_{i_2}\ldots x_{i_{2k}}\right).$$

This function has degree $m-1$ and, omitting terms of degree less than $m-2$, is trivially EL^{m-3}–equivalent to the following function which we will use as f_1^* from now on:

$$f_1^* = \overline{x_m} + x_m f^\star,$$

where $f^\star$, defined by

$$f^\star(x_1, x_2, \ldots, x_{m-1}) = \left(\sum_{1 \leqslant i < j \leqslant m-1} \overline{x_i x_j} \right)$$

is a function of the first $m-1$ variables. Moving on we will denote the $(m-1)$-tuple of the first $m-1$ variables as $\bar{\mathrm{x}}$. We will also denote affine transforms of the first $m-1$ variables as $\bar{L}_{\mathrm{A}}^{\mathrm{b}}$ (with matrix and vector of corresponding sizes).

Let us now give an explicit description of a certain function from F_2. Denote as f_2^* the function with the support $\{0, \mathrm{e}_1, \mathrm{e}_2, \ldots, \mathrm{e}_{m-1}, \sum_{i=1}^{m-1} \mathrm{e}_i\}$. After a bit of calculation one can explicitly describe its ANF:

$$f_2^*(\mathrm{x}) = (1 + x_m) \left(1 + \sum_{k=1}^{\frac{m-3}{2}} \sum_{1 \leqslant i_1 < \ldots < i_{2k} \leqslant m-1} x_{i_1} x_{i_2} \ldots x_{i_{2k}} \right).$$

This function has degree $m-1$ and is trivially EL^{m-3} equivalent to the function $x_m f^\star$, which we will use as f_2^* from now on.

Note that $f_1^* = \overline{x_m} + f_2^*$.

Let us build some alternative representatives of the equivalence classes of F_1 and F_2. The following lemma will be helpful:

Lemma 7 *Let $f = \overline{x_m} + h$, where $\deg(h) \leqslant m-2$. Let A be a nonsingular $m \times m$ matrix. Then $f \circ L_{\mathrm{A}} = \overline{x_m} + h_1$ for some h_1 of degree at most $m-2$ if and only if matrix A has the following form:*

$$\mathrm{A} = \begin{pmatrix} \bar{\mathrm{A}} & 0^{m-1} \\ \mathrm{w} & 1 \end{pmatrix},$$

where 0^{m-1} is an all-zero column of length $m-1$, $\bar{\mathrm{A}}$ is an arbitrary nonsingular $(m-1) \times (m-1)$ matrix and w is an arbitrary row of length $m-1$.

Proof. See Appendix III. □

This lemma shows us that all linear transformations of the described form, and only such transformations among all linear, transform functions of the form $\overline{x_m} + h$ with $\deg(h) < m-1$ into functions of the same form, preserving $\overline{x_m}$ as the only monomial of degree $m-1$.

Let g be an arbitrary function of degree $m-1$ with the highest-degree component of the form $\overline{x_m}$:

$$g = \overline{x_m} + x_m g_1 + g_2,$$

where g_1, g_2 do not depend on x_m and $\deg(g_1) < m-2$, $\deg(g_2) < m-1$. Let us look at the action of a transformation described in the Lemma 7 onto such function

g. Assume that A is a matrix satisfying conditions of Lemma 7. Discarding terms of degree less than $m-2$, we have:

$$g \circ L_{\mathrm{A}} = \overline{x_m} + x_m(g_1 \circ \bar{L}_{\bar{\mathrm{A}}}) + g_3,$$

where g_3 is some function of degree at most $m-2$ which doesn't depend on the variable x_m.

Let us now build alternative representatives for the metric complement of $\mathcal{R}_{m-3}$. Since $\bar{\mathrm{A}}$ can be an arbitrary nonsingular matrix, choosing $\bar{\mathrm{A}}$ so that $f^\star \circ \bar{L}_{\bar{\mathrm{A}}} = p_{\frac{m-1}{2}}$ (Lemma 5) and filling w with zeroes, we can obtain a matrix A such that

$$f^*_{1\mathrm{A}} := f^*_1 \circ L_{\mathrm{A}} = \overline{x_m} + x_m(f^\star \circ \bar{L}_{\bar{\mathrm{A}}}) + h_1 = \overline{x_m} + x_m p_{\frac{m-1}{2}} + h_1.$$

Here $p_{\frac{m-1}{2}}, h_1$ do not depend on x_m and h_1 has degree at most $m-2$. Additionally,

$$f^*_{2\mathrm{A}} := f^*_2 \circ L_{\mathrm{A}} = x_m(f^\star \circ \bar{L}_{\bar{\mathrm{A}}}) = x_m p_{\frac{m-1}{2}}.$$

We will use these equivalent functions $f^*_{1\mathrm{A}}$ and $f^*_{2\mathrm{A}}$ as class representatives in some cases.

8.3 **Metric regularity**

We have established that

$$\widehat{\mathcal{R}}_{m-3} = \{g : g \overset{m-3}{\sim} f^*_1\} \cup \{g : g \overset{m-3}{\sim} f^*_2\},$$

where f^*_1 is a representative of an EL^{m-3}–equivalence class of F_1 and f^*_2 is a representative of an EL^{m-3}–equivalence class of F_2, and have presented some variants of these representatives.

Since the code $\mathcal{R}_{m-3}$ is linear, it follows [8] that $\rho(\widehat{\mathcal{R}}_{m-3}) = \rho(\mathcal{R}_{m-3}) = m+2$ and function f is in $\widehat{\widehat{\mathcal{R}}}_{m-3}$ if and only if $f + \widehat{\mathcal{R}}_{m-3} = \widehat{\mathcal{R}}_{m-3}$. With this in mind, let us prove the metric regularity of $\mathcal{R}_{m-3}$ by proving that no functions other that those contained in $\mathcal{R}_{m-3}$ preserve the metric complement under addition.

Case 1. Let $f \notin \mathcal{R}_{m-3}$ be a function of degree greater than $m-1$. Since EL^{m-3}–equivalence preserves degree of functions with degree higher than $m-3$, any $g \in \widehat{\mathcal{R}}_{m-3}$ has degree $m-2$ or $m-1$, while $f+g$ has higher degree and therefore cannot be equivalent to any of the functions from $\widehat{\mathcal{R}}_{m-3}$. Thus, functions of degree greater than $m-1$ cannot be in $\widehat{\widehat{\mathcal{R}}}_{m-3}$.

Case 2. Let $f \notin \mathcal{R}_{m-3}$ be a function of degree $m-1$. Any function of degree $m-1$ is trivially EL^{m-3}–equivalent to some function with $\overline{x_m}$ as the only monomial of degree $(m-1)$, so

$$f \circ L_{\mathrm{B}} + h = \overline{x_m} + x_m f_1 + f_2, \tag{1}$$

for some nonsingular B and function h of degree at most $m-3$. Here f_1, f_2 do not depend on x_m, f_1 is either zero or has degree $m-3$, while f_2 is either zero or has degree $m-2$.

Case 2.1. If f_1 in (1) is nonzero, then, using Lemma 7 and Lemma 5, we can pick such a matrix B and function h of degree at most $m-3$ that

$$f \circ L_{\mathrm{B}} + h = \overline{x_m} + x_m p_k + f_3$$

for some $k > 0$ and some f_3 of degree at most $m-2$ (p_k, f_3 do not depend on x_m). If we now sum f and $(f^*_{2\mathrm{A}} + h) \circ L_{\mathrm{B}^{-1}} \in \widehat{\mathcal{R}}_{m-3}$, we obtain a function $f^\dagger$, which is equivalent to:

$$f^\dagger \overset{m-3}{\sim} f^*_{2\mathrm{A}} + f \circ L_{\mathrm{B}} + h = \overline{x_m} + x_m(p_k + p_{\frac{m-1}{2}}) + f_3 \overset{m-3}{\sim} \overline{x_m} + x_m p_{\frac{m-1}{2}-k} + f_4,$$

where f_4 is some function of degree at most $m-2$, not depending on x_m, and the last equivalence is just variable renaming.

Let us denote the function on the right of (1) as $f^\sim$. It has degree $m-1$ and therefore cannot be equivalent to f^*_2. It cannot be equivalent to $f^*_{1\mathrm{A}}$ either, because, by Lemma 7, any linear transformation of variables with matrix D preserving $\overline{x_m}$, will act onto it in the following manner:

$$f^\sim \circ L_{\mathrm{D}} = \overline{x_m} + x_m(p_{\frac{m-1}{2}-k} \circ \bar{L}_{\bar{\mathrm{D}}}) + f_5,$$

where f_5 is some function of degree at most $m-2$ in the first $m-1$ variables. It is clear that no matrix $\bar{\mathrm{D}}$ can match the $(m-2)$-degree parts of $f^\sim$ and $f^*_{1\mathrm{A}}$ containing variable x_m, since $p_{\frac{m-1}{2}-k}$ is not equivalent to $p_{\frac{m-1}{2}}$. Thus, $f^\dagger = f + (f^*_{2\mathrm{A}} + h) \circ L_{\mathrm{B}^{-1}}$ is not in $\widehat{\mathcal{R}}_{m-3}$, and therefore, if f_1 is nonzero, f is not in $\widehat{\widehat{\mathcal{R}}}_{m-3}$.

Case 2.2. If f_1 in (1) is equal to zero, we have a couple more cases to study.

Assume that both f_1 and f_2 are zero and thus

$$f \circ L_{\mathrm{B}} + h = \overline{x_m}.$$

Using transformation $x_1 \to x_1 + x_m$ (and removing terms of degree less than $m-2$), the function $f^*_1 = \overline{x_m} + x_m f^\star$ transforms into $\overline{x_m} + \overline{x_1} + x_m f^\star$.

If we now add f and $(\overline{x_m} + \overline{x_1} + x_m f^\star + h) \circ L_{\mathrm{B}^{-1}} \in \widehat{\mathcal{R}}_{m-3}$ we will obtain a function $f^\dagger$, which is equivalent to:

$$f^\dagger \overset{m-3}{\sim} \overline{x_m} + \overline{x_1} + x_m f^\star + f \circ L_{\mathrm{B}} + h = \overline{x_1} + x_m f^\star.$$

If we swap the variables x_1 and x_m in the right-hand side by another linear transformation and regroup terms, we will obtain the following function:

$$\overline{x_m} + \sum_{2 \leqslant i < j \leqslant m-1} \overline{x_i x_j} + \sum_{i=2}^{m-1} \overline{x_i x_m},$$

which is in turn equivalent to

$$\overline{x_m} + x_m p_{\frac{m-3}{2}} + h^\dagger$$

for some $h^\dagger$ of degree at most $m-2$ in the first $m-1$ variables. By Lemma 7 and Lemma 5, this function cannot be equivalent to $f^*_{1\mathrm{A}}$ and it is not equivalent to f^*_2 by degree comparison. Thus, $f^\dagger = f + (\overline{x_m} + \overline{x_1} + x_m f^\star + h) \circ L^{-1}_{1m}$ is not in $\widehat{\mathcal{R}}_{m-3}$, and therefore f is not in $\widehat{\widehat{\mathcal{R}}}_{m-3}$.

Case 2.3. Assume that f_1 is zero and f_2 is nonzero in (1). Then

$$f \circ L_{\mathrm{B}} + h = \overline{x_m} + f_2.$$

Since f_2 doesn't contain the variable x_m, all terms of f_2 are of the form $\overline{x_i x_m}$ for some i. Without loss of generality (swapping variables among first $m-1$ if needed) we can assume that f_2 contains $\overline{x_{m-1}x_m}$. Renaming variables in $f^*_{2\mathrm{A}}$, we can transform it into:

$$\overline{x_2 x_3} + \overline{x_4 x_5} + \ldots + \overline{x_{m-1}x_m}.$$

If we now add f and the above function, which belongs to $\widehat{\mathcal{R}}_{m-3}$ we will obtain a function $f^\dagger$, which is equivalent to:

$$f^\dagger \overset{m-3}{\sim} \sum_{k=1}^{\frac{m-1}{2}} \overline{x_{2k}x_{2k+1}} + f \circ L_{\mathrm{B}} + h = \overline{x_m} + \sum_{k=1}^{\frac{m-3}{2}} \overline{x_{2k}x_{2k+1}} + \sum_{i \in I} \overline{x_i x_m}$$

which is equivalent to

$$\overline{x_m} + x_m p_{\frac{m-3}{2}} + h^\dagger$$

for some $h^\dagger$ of degree at most $m-2$ in the first $m-1$ variables. By Lemma 7 and Lemma 5, this function cannot be equivalent to $f^*_{1\mathrm{A}}$ and it is not equivalent to f^*_2 by degree comparison. Thus, $f^\dagger = f + \left(\sum_{k=1}^{\frac{m-1}{2}} \overline{x_{2k}x_{2k+1}} + h \right) \circ L_{\mathrm{B}^{-1}}$ is not in $\widehat{\mathcal{R}}_{m-3}$, and therefore f is not in $\widehat{\widehat{\mathcal{R}}}_{m-3}$.

Case 3. If $f \notin \mathcal{R}_{m-3}$ is a function of degree $m-2$, then, by arguments similar to the case of even m, f is equivalent to p_k (in m variables) for some $k > 0$ using some nonsingular linear transform L and some addition h of degree at most $m-3$. Then

$$f \circ L + h + f^*_{2\mathrm{A}} \overset{m-3}{\sim} p_{\frac{m-1}{2}-k},$$

and therefore $g = h \circ L^{-1} + f^*_{2\mathrm{A}} \circ L^{-1}$ is the function from the metric complement such that $f+g$ is inequivalent to both $f^*_{2\mathrm{A}}$ (because $\frac{m-1}{2} \neq \frac{m-1}{2} - k$) and f^*_1 (by degree comparison).

Since all functions which are not in $\mathcal{R}_{m-3}$ have degree $m-2$ or higher, we have proven that none of them are in the double metric complement, and therefore $\mathcal{R}_{m-3}$ is metrically regular.

9 Conclusion

In this paper we have established metric regularity of the $\mathcal{RM}(1,5)$ code and of the codes $\mathcal{RM}(k,m)$ for $k \geqslant m-3$. Factoring in the result by Tokareva [14], which proves metric regularity of $\mathcal{RM}(1,m)$ for even m, we have covered all infinite families of Reed-Muller codes with known covering radius. The only other Reed-Muller codes with known covering radius, metric regularity of which has not been yet established, are $\mathcal{RM}(1,7)$, $\mathcal{RM}(2,6)$ and $\mathcal{RM}(2,7)$. Given these results, we formulate the following

Conjecture. *All Reed-Muller codes $\mathcal{RM}(k,m)$ are metrically regular.*

References

1. Berlekamp E., Welch L.: Weight distributions of the cosets of the $(32, 6)$ Reed-Muller code. IEEE Transactions on Information Theory. **18**(1), 203–207 (1972).

2. Cohen, G., Honkala, I., Litsyn, S., Lobstein, A.: Covering codes. Elsevier. **54**, (1997).

3. Hou X. D.: Covering Radius of the Reed-Muller code $R(1, 7)$ – A Simpler Proof. Journal of Combinatorial Theory, Series A. **74**(2), 337–341 (1996).

4. Kutsenko, A.: Metrical properties of self-dual bent functions. Designs, Codes and Cryptography (2019). doi:10.1007/s10623-019-00678-x

5. McLoughlin A. M.: The Covering Radius of the $(m-3)$-rd Order Reed Muller Codes and a Lower Bound on the $(m-4)$-th Order Reed Muller Codes. SIAM Journal on Applied Mathematics. **37**(2), 419–422 (1979).

6. Mykkeltveit J.: The covering radius of the $(128, 8)$ Reed-Muller code is 56. IEEE Transactions on Information Theory. **26**(3), 359–362 (1980).

7. Neumaier A.: Completely regular codes. Discrete mathematics. **106**, 353–360 (1992).

8. Oblaukhov A. K.: Metric complements to subspaces in the Boolean cube. Journal of Applied and Industrial Mathematics. **10**(3), 397–403 (2016).

9. Oblaukhov A. K.: Maximal metrically regular sets. Siberian Electronic Mathematical Reports. **15**, 1842–1849 (2018).

10. Oblaukhov A.: A lower bound on the size of the largest metrically regular subset of the Boolean cube. Cryptography and Communications. **11**(4), 777–791 (2019).

11. Rothaus O. S.: On "bent" functions. Journal of Combinatorial Theory, Series A. **20**(3), 300–305 (1976).

12. Schatz J.: The second order Reed-Muller code of length 64 has covering radius 18. IEEE Transactions on Information Theory. **27**(4), 529–530 (1981).

13. Stanica P., Sasao T., Butler J. T.: Distance duality on some classes of Boolean functions. Journal of Combinatorial Mathematics and Combinatorial Computing. 2018.

14. Tokareva N.: Duality between bent functions and affine functions. Discrete Mathematics. **312**(3), 666–670 (2012).

15. Tokareva N.: Bent functions: results and applications to cryptography. Academic Press, (2015).

16. Wang Q.: The covering radius of the Reed–Muller code $RM(2, 7)$ is 40. Discrete Mathematics. **342**(12), Article 111625 (2019).

No	Representative f	Added $g \in \widehat{\mathcal{RM}}(1,5)$	$C(g)$	Sum $h = f + g$	$C(h)$
0	0	—	—	—	—
1	2345	123+14+25	22	2345+123+14+25	12
2	2345+14	123+14+25	22	2345+123+25 ∼ 2345+123+34	8
3	2345+24	2345+123+24+35	14	123+35 ∼ 123+14	21
4	2345+24+35	2345+123+24+35	14	123	19
5	2345+14+25	123+14+25	22	2345+123	6
6	2345+123	123+14+25	22	2345+14+25	5
7	2345+123+12	12+34	28	2345+123+34	8
8	2345+123+34	12+34	28	2345+123+12	7
9	2345+123+14	14+25	28	2345+123+25 ∼ 2345+123+34	8
10	2345+123+45	12+45	28	2345+123+12	7
11	2345+123+12+34	12+34	28	2345+123	6
12	2345+123+14+25	123+14+25	22	2345	1
13	2345+123+12+45	12+45	28	2345+123	6
14*	2345+123+24+35	2345+123+24+35	14	0	0
15	2345+123+145	123+14+25	22	2345+145+14+25 ∼ 2345+123+12+34	11
16	2345+123+145+45	123+145+45+24+35	26	2345+24+35	4
17	2345+123+145+24+45	2345+123+24+35	14	145+35+45 ∼ 123+14	21
18	2345+123+145+24+35	2345+123+24+35	14	145 ∼ 123	19
19	123	2345+123+24+35	14	2345+24+35	4
20	123+45	2345+123+24+35	14	2345+24+35+45 ∼ 2345+24+35	4
21	123+14	123+14+25	22	25 ∼ 12	27
22*	123+14+25	123+14+25	22	0	0
23	123+145	123+14+25	22	145+14+25 ∼ 145+25 ∼ 123+14	21
24	123+145+23	23+45	28	123+145+45 ∼ 123+145+23	24
25	123+145+24	123+15+24	22	145+15 ∼ 123	19
26*	123+145+45+24+35	123+145+45+24+35	26	0	0
27	12	12+34	28	34 ∼ 12	27
28*	12+34	12+34	28	0	0

Table 1 Table of even weight cosets of $\mathcal{RM}(1,5)$ [1]. Classes marked with an asterisk are those which constitute $\widehat{\mathcal{RM}}(1,5)$. $C(\cdot)$ denotes the No of the class the function belongs to.

Appendix I: Metric regularity of $\mathcal{RM}(1,5)$

Theorem 1 *The code $\mathcal{RM}(1,5)$ is metrically regular.*

Proof. It is known [1] that the cosets of the $\mathcal{RM}(1,5)$ code can be partitioned into 48 classes with respect to the EA-equivalence. Four of these coset classes contain coset leaders with the largest attainable weight 12 (classes 14, 22, 26 and 28 in the Table 1), which proves

$$\rho(\mathcal{RM}(1,5)) = 12.$$

Since $\rho(\widehat{\mathcal{RM}}(1,5)) = \rho(\mathcal{RM}(1,5)) = 12$, the second metric complement of $\mathcal{RM}(1,5)$ can consist only of the cosets with codewords of even weight. There are 29 classes of such cosets, including the $\mathcal{RM}(1,5)$ code itself; they are listed in the Table 1. Classes marked with an asterisk are those which constitute $\widehat{\mathcal{RM}}(1,5)$. These classes were obtained in the work [1] by Berlekamp and Welch. In this work some of the class representatives were modified from their original variants using simple variable swaps. Functions in the table are presented in an abbreviated notation: the number $i_1 i_2 \dots i_k$ stands for the monomial $x_{i_1} x_{i_2} \dots x_{i_k}$. For example, the representative function for the class 14 is $x_2x_3x_4x_5 + x_1x_2x_3 + x_2x_4 + x_3x_5$.

As has been shown in the Section 3, in order to prove that no codeword from a certain coset class C belongs to the second metric complement $\widehat{\widehat{\mathcal{RM}}}(1,5)$, we must prove that $f + g \notin \widehat{\mathcal{RM}}(1,5)$ for some $f \in C$ and some $g \in \widehat{\mathcal{RM}}(1,5)$. The proof can be found in the Table 1: for a representative f from each even weight coset class we find a function $g \in \widehat{\mathcal{RM}}(1,5)$ such that $f + g$ is equivalent

to the representative of some class which is not in $\widehat{\mathcal{RM}}(1,5)$. Thus, the second metric complement $\widehat{\widehat{\mathcal{RM}}}(1,5)$ contains only the code $\mathcal{RM}(1,5)$ itself, proving that $\mathcal{RM}(1,5)$ is metrically regular.

Almost all equivalences presented in the fifth column of the table are variable swaps or additions of the form $x_i \to x_i + 1$, $x_i \to x_i + x_j$ or (for the class 20) $x_i \to x_i + x_j + x_k$.

□

Appendix II: Minimal syndrome matrices

Let us remember Lemma 1, since it is extensively used when proving subsequent lemmas.

Lemma 1 *Let* $\mathbf{S}$ *be a symmetric matrix, and let* $\mathbf{B}$ *be its factor (i.e.* $\mathbf{BB^T} = \mathbf{S}$*). The following operations do not change the property of* $\mathbf{B}$ *being a factor of* $\mathbf{S}$*:*

1. *deleting a zero column;*
2. *deleting two equal columns;*
3. *swapping any two columns;*
4. *adding an arbitrary vector b to each column from some subset of columns of* $\mathbf{B}$ *of even size, given that all columns of this subset sum to zero.*

Lemma 2 *Let* $\mathbf{S}$ *be a symmetric matrix, and let* $\mathbf{B}$ *be its minimal factor. Then all proper subsets of columns of* $\mathbf{B}$ *are linearly independent.*

Proof. Assume that $\mathbf{B}$ has a proper linearly dependent subset of columns which sum to zero. If this subset has odd number of columns, we make it even by adding a zero vector to it (and to the matrix $\mathbf{B}$).

Let us denote the matrix comprised of the vectors from this subset as $\bar{\mathbf{B}}$. Then, swapping columns if required, the matrix $\mathbf{B}$ can be represented as $\mathbf{B} = (\bar{\mathbf{B}}, \mathbf{D})$, where $\bar{\mathbf{B}}$ has even number of columns summing to zero, while $\mathbf{D}$ is a nonempty matrix, consisting of all remaining columns.

Let $\mathbf{b^1}$ and $\mathbf{d^1}$ be the first (nonzero) columns of $\bar{\mathbf{B}}$ and $\mathbf{D}$ respectively. Let us add the column $\mathbf{b^1} + \mathbf{d^1}$ to all columns of $\bar{\mathbf{B}}$ — we can do that by Lemma 1, without changing the property of $\mathbf{B}$ being a factor of $\mathbf{S}$. Now the first (nonzero) columns of $\bar{\mathbf{B}}$ and $\mathbf{D}$ are equal, and we can delete them by Lemma 1.

Thus, we have added not more that one zero column to the factor $\mathbf{B}$ and then removed two columns from it. This decreases the number of columns in $\mathbf{B}$, which contradicts with its minimality. □

Lemma 3 *Let* $\mathbf{S}$ *be a symmetric* $m \times m$ *matrix, where* m *is even. Then* $t(\mathbf{S}) = m + 1$ *if and only if* $\mathbf{S} = \mathbf{BB^T}$ *for some* $m \times (m+1)$ *matrix* $\mathbf{B}$ *of rank* m *such that all its columns sum to zero.*

Proof. $\Longrightarrow$

Assume that $t(\mathbf{S}) = m + 1$, and $\mathbf{B}$ is a minimal factor of $\mathbf{S}$.

If some m-subset of columns of $\mathbf{B}$ is not linearly independent, then, by Lemma 2, $\mathbf{B}$ cannot be a minimal factor of $\mathbf{S}$. Thus, each m-subset of columns of $\mathbf{B}$ is linearly independent and $\mathbf{B}$ has rank m.

Since $\mathbf{B}$ has $m+1$ columns of length m, some subset of them must sum to zero. If all columns of $\mathbf{B}$ do not sum to zero, then it must be a proper subset, which contradicts with the minimality of $\mathbf{B}$ by Lemma 2. Therefore all columns of $\mathbf{B}$ sum to zero.

$\Longleftarrow$

Assume that $\mathbf{S} = \mathbf{B}\mathbf{B}^\mathbf{T}$, where $\mathbf{B}$ is a $m \times (m+1)$ matrix of rank m with of all its columns summing to $\mathbf{0}$. It is easy to see that each m-subset of columns of such $\mathbf{B}$ is, in fact, linearly independent. Let us prove that $\mathbf{B}$ is a minimal factor of $\mathbf{S}$.

Assume that $t(\mathbf{S}) = k \leqslant m$ and $\mathbf{D}$ is an arbitrary minimal factor of $\mathbf{S}$ with k columns. Since the sum of all columns of a factor is the vector consisting of the diagonal elements of $\mathbf{S}$, sum of all columns of $\mathbf{D}$ is also equal to zero. This implies that $\text{rank}(\mathbf{D}) < m$, and therefore $\text{rank}(\mathbf{S}) < m$, since $\mathbf{S} = \mathbf{D}\mathbf{D}^\mathbf{T}$.

This means that there exists a subset of rows in $\mathbf{S}$ summing to $\mathbf{0}$; we denote these rows as $\mathbf{S_{i_1}}, \mathbf{S_{i_2}}, \ldots, \mathbf{S_{i_p}}$. Since $\mathbf{S_i} = \mathbf{B_i}\mathbf{B}^\mathbf{T}$, this implies

$$(\mathbf{B_{i_1}} + \ldots + \mathbf{B_{i_p}})\mathbf{B}^\mathbf{T} = \mathbf{0}.$$

Denote $\mathbf{b} = \mathbf{B_{i_1}} + \ldots + \mathbf{B_{i_p}}$. From the above it follows that the sum of certain columns of $\mathbf{B}$ (in particular, the columns corresponding to 1's in the vector $\mathbf{b}$) is equal to zero.

If the vector $\mathbf{b}$ is zero, then rows of $\mathbf{B}$ are not linearly independent and $\text{rank}(\mathbf{B}) < m$, contradiction.

If it is nonzero and not an all-ones vector, then we obtain a proper subset of columns of $\mathbf{B}$ which sum to 0, contradiction.

If $\mathbf{b}$ is an all-ones vector, then, since m is even, the length of the vector $\mathbf{b}$ is odd and therefore $\mathbf{b}\mathbf{b}^\mathbf{T} = 1$, which contradicts with $(\mathbf{B_1} + \ldots + \mathbf{B_k})\mathbf{B}^\mathbf{T} = \mathbf{0}$.

Thus, $t(\mathbf{S}) = m+1$ and $\mathbf{B}$ is a minimal factor of $\mathbf{S}$.

□

Proposition 1 *Let* $\mathbf{S}$ *be a symmetric matrix. Then* $t(\mathbf{S}) = m+1$ *if and only if* $\text{rank}(\mathbf{S}) = m$ *and* $\mathbf{S}$ *has an all-zero diagonal.*

Proof. $\Longrightarrow$

Let $\mathbf{S}$ be a symmetric matrix with $t(\mathbf{S}) = m+1$, and let $\mathbf{B}$ be any of its minimal factors. Assume that $\text{rank}(\mathbf{S}) < m$. Then there exists a subset of rows in $\mathbf{S}$ summing to $\mathbf{0}$; we denote these rows as $\mathbf{S_{i_1}}, \mathbf{S_{i_2}}, \ldots, \mathbf{S_{i_p}}$. Since $\mathbf{S_i} = \mathbf{B_i}\mathbf{B}^\mathbf{T}$, this implies

$$(\mathbf{B_{i_1}} + \ldots + \mathbf{B_{i_p}})\mathbf{B}^\mathbf{T} = \mathbf{0}.$$

Denote $\mathbf{b} = \mathbf{B_{i_1}} + \ldots + \mathbf{B_{i_p}}$. From the above it follows that the sum of certain columns of $\mathbf{B}$ (in particular, the columns corresponding to ones in the vector $\mathbf{b}$) is equal to zero.

If the vector $\mathbf{b}$ is zero, then $\text{rank}(\mathbf{B}) < m$ and it must have a linearly dependent proper subset of columns, hence $\mathbf{B}$ is not minimal by Lemma 2.

If it is nonzero and not an all-ones vector, then we obtain a proper subset of columns of $\mathbf{B}$ which sum to 0, hence $\mathbf{B}$ is not minimal by Lemma 2.

If $\mathbf{b}$ is an all-ones vector, then all columns of $\mathbf{B}$ sum to zero.

If m is even, then the number of columns in $\mathbf{B}$ is odd and therefore $\mathbf{b}\mathbf{b}^\mathbf{T} = 1$, which contradicts with $(\mathbf{B_1} + \ldots + \mathbf{B_k})\mathbf{B}^\mathbf{T} = \mathbf{0}$.

If m is odd, then the number of columns in $\mathbf{B}$ is even and all rows have even number of ones, so by Lemma 1, we can add any column of $\mathbf{B}$ to all its columns and then delete a zero column from the result, keeping it a factor of $\mathbf{S}$, which contradicts the minimality of $\mathbf{B}$.

Thus, $\mathrm{rank}(\mathbf{S}) = m$.

Assume that $\mathbf{S}$ has a non-zero diagonal. Since the vector consisting of diagonal elements of $\mathbf{S}$ is the sum of all columns of $\mathbf{B}$, all columns of $\mathbf{B}$ sum to a non-zero vector, which means that $\mathbf{B}$ must have a proper subset of columns summing to $\mathbf{0}$, thus contradicting the minimality of $\mathbf{B}$.

$\Longleftarrow$

Clearly, $t(\mathbf{S}) \geqslant \mathrm{rank}(\mathbf{S}) = m$. Let $\mathbf{B}$ be a minimal factor of $\mathbf{S}$. Trivially, the rank of $\mathbf{B}$ is also equal to m. Since the diagonal of $\mathbf{S}$ is all-zero, all columns of $\mathbf{B}$ sum to zero, hence it cannot have only m columns while having rank m, and must have at least $m + 1$. □

Lemma 6 *Let $\mathbf{S}$ be a symmetric $m \times m$ matrix, where m is odd. Then $t(\mathbf{S}) \leqslant m$, and the equality is achieved if and only if $\mathbf{S} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$ for some $m \times m$ matrix $\mathbf{B}$ which is either nonsingular, or has rank $m - 1$ and all columns summing to zero.*

Proof. As mentioned in the Section 6, $t(\mathbf{S})$ is at most $m + 1$ for any $\mathbf{S}$. Assume that $t(\mathbf{S}) = m + 1$. By Proposition 1, this can only happen if $\mathrm{rank}(\mathbf{S}) = m$ and $\mathbf{S}$ has all-zero diagonal, which is impossible in the case of odd m (see e.g. [2] p. 249). Thus $t(\mathbf{S}) \leqslant m$.

$\Longrightarrow$

Assume that $t(\mathbf{S}) = m$ and let $\mathbf{B}$ be a minimal factor of $\mathbf{S}$ with m columns. If the rank of $\mathbf{B}$ is smaller than $m - 1$, then $\mathbf{B}$ has a proper subset of columns summing to zero, contradicting minimality of $\mathbf{B}$, so the rank of a factor must be at least $m - 1$. If the rank is m, the proof is finished.

Assume that $\mathrm{rank}(\mathbf{B}) = m - 1$. Then some subset of columns of $\mathbf{B}$ must sum to zero. Since $\mathbf{B}$ is minimal, it cannot be a proper subset by Lemma 2, therefore all columns of $\mathbf{B}$ must sum to zero.

$\Longleftarrow$

Clearly, $t(\mathbf{S}) \geqslant \mathrm{rank}(\mathbf{S})$, so if $\mathbf{S} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$ for some nonsingular $m \times m$ matrix $\mathbf{B}$, then the proof is finished.

Let $\mathbf{S} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$ for some $\mathbf{B}$ of rank $m - 1$ with all columns summing to zero.

Assume that $t(\mathbf{S}) = k \leqslant m-1$ and let $\mathbf{D}$ be some minimal factor of $\mathbf{S}$. Note that the sum of all columns of any factor is the vector composed of diagonal elements of $\mathbf{S}$, so the sum of all columns of $\mathbf{D}$ is also zero.

Assume that $k = m - 1$. Then $\mathbf{D}$ has even number of columns, and each row has even number of ones, so we can add an arbitrary vector to all columns of $\mathbf{D}$ while keeping it a factor of $\mathbf{S}$ using Lemma 1. Let us add the first column of $\mathbf{D}$ to all columns of $\mathbf{D}$. Now the first column of $\mathbf{D}$ is zero and we can remove it by Lemma 1. We have now obtained a factor of $\mathbf{S}$ with fewer columns than in $\mathbf{D}$, which contradicts with the minimality of $\mathbf{D}$.

Therefore, k must be at most $m-2$. As mentioned before, the sum of all columns of $\mathbf{D}$ is zero, which means that $\mathbf{D}$ is not a full-rank matrix. Hence $\mathrm{rank}(\mathbf{D})$ is at most $m - 3$, which means that $\mathrm{rank}(\mathbf{S})$ is at most $m - 3$ as well.

Since $\mathbf{S} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$, by Sylvester's inequality we obtain $\text{rank}(\mathbf{S}) \geqslant \text{rank}(\mathbf{B}) + \text{rank}(\mathbf{B}^{\mathbf{T}}) - m = m - 2$. But we have just proved that $\text{rank}(\mathbf{S}) \leqslant m - 3$, contradiction. Thus $t(\mathbf{S})$ must be greater than $m - 1$ and is equal to m.

□

Appendix III: Other results

Lemma 4 *Let C be a code with the covering radius r and the metric complement $\widehat{C}$. Let C_π be the code obtained from C by adding a parity check bit. Then $\rho(C_\pi) = r + 1$ and $\widehat{C}_\pi$ is obtained from $\widehat{C}$ by*

1. *adding a parity check bit to all vectors in case if r is odd or*
2. *adding a reverse parity check bit to all vectors in case if r is even.*

Proof. Assume that r is even. Denote

$$C_0 = \{c \in C : wt(c) \text{ is even}\},\ C_1 = \{c \in C : wt(c) \text{ is odd}\},$$

$$\widehat{C}_0 = \{c \in \widehat{C} : wt(c) \text{ is even}\},\ \widehat{C}_1 = \{c \in \widehat{C} : wt(c) \text{ is odd}\}.$$

Due to r being even, vectors from $\widehat{C}_0$ are at distance r from C_0 and at a larger distance from C_1. Similarly, vectors from $\widehat{C}_1$ are at distance r from C_1 and at a larger distance from C_0. Denote

$$C_{\pi,0} = \{(0, c) : c \in C_0\},\ C_{\pi,1} = \{(1, c) : c \in C_1\}.$$

Clearly, $C_\pi = C_{\pi,0} \cup C_{\pi,1}$.

Let $v = (1, c)$, where $c \in \widehat{C}_1$. It is easy to see that $d(v, C_{\pi,1}) = d(c, C_1) = r$. Let $v = (0, c)$, where $c \in \widehat{C}_1$. It follows that $d(v, C_{\pi,1}) = d(c, C_1) + 1 = r + 1$ and $d(v, C_{\pi,0}) = d(c, C_0) \geqslant r + 1$.

Let $v = (0, c)$, where $c \in \widehat{C}_0$. It follows that $d(v, C_{\pi,0}) = d(c, C_0) = r$. Let $v = (1, c)$, where $c \in \widehat{C}_0$. It follows that $d(v, C_{\pi,0}) = d(c, C_0) + 1 = r + 1$ and $d(v, C_{\pi,1}) = d(c, C_1) \geqslant r + 1$.

Let $v = (\epsilon, c)$, where $c \notin \widehat{C}$ and $\epsilon \in \{0, 1\}$. It follows that $d(v, C_\pi) \leqslant d(c, C) + 1 \leqslant r$.

We have examined all possibilities for the vector v and the claim of the Lemma follows from these examinations.

The case of odd r is similar. □

Lemma 5 *Let f and g be two function of degree $m-2$. Then $f \overset{m-3}{\sim} g$ if and only if their quadratic duals are* EL^1*-equivalent (*EA*-equivalent).*

Proof. Since EL^{m-3}-equivalence allows us to add functions of degree up to $m-3$, we will assume that both f and g contain only monomials of degree $m - 2$. In what follows we will discard monomials of degree less than $m - 2$ when talking about EL^{m-3}-equivalence, and we will discard monomials of degree less than 2 when talking about EL^1-equivalence.

Let $f(\mathrm{x}) = \sum\limits_{(i,j)\in I} \overline{x_i x_j}$ be the ANF of f. Let us perform the following simple nonsingular linear transformation of variables L_{ij}:

$$L_{ij} : \begin{cases} x_i \to x_i + x_j, \\ x_k \to x_k \qquad\quad \forall k \neq i. \end{cases}$$

Let us inspect how the function f changes under this transformations (disregarding monomials of degree less than $m-2$):

$$L_{ij} : \begin{cases} \overline{x_i x_k} \to \overline{x_i x_k} & \forall k \neq i, \\ \overline{x_j x_k} \to \overline{x_j x_k} + \overline{x_i x_k} & \forall k \neq i, j, \\ \overline{x_k x_l} \to \overline{x_k x_l} & \forall k, l \neq i, j. \end{cases}$$

Denote the function obtained after this transformation as f_1. Then it is easy to see that the dual $\tilde{f}_1$ is obtained from the dual $\tilde{f}$ (disregarding monomials of degree less than 2 since we consider EL^1-equivalence) by the following linear transformation:

$$\begin{cases} x_j \to x_j + x_i, \\ x_k \to x_k \qquad\quad \forall k \neq j. \end{cases}$$

which is simply the transposed transformation L_{ji}.

Assume now that g is obtained from f using linear transformation L. Then L can be decomposed into a sequence of simple transformations:

$$L = L_{i_1 j_1} \circ L_{i_2 j_2} \circ \ldots \circ L_{i_s j_s}.$$

From the above we can see that the dual $\tilde{g}$ is obtained from $\tilde{f}$ using the following transformation $\tilde{L}$:

$$\tilde{L} = L_{j_1 i_1} \circ L_{j_2 i_2} \circ \ldots \circ L_{j_s i_s}$$

which is a sequence of transposed simple transformations.

Thus, if $f \overset{m-3}{\sim} g$, then $\tilde{f} \overset{1}{\sim} \tilde{g}$. The reverse can be proven using similar arguments.

□

Lemma 7 *Let $f = \overline{x_m} + h$, where $\deg(h) \leqslant m-2$. Let A be a nonsingular $m \times m$ matrix. Then $f \circ L_{\mathrm{A}} = \overline{x_m} + h_1$ for some h_1 of degree at most $m-2$ if and only if matrix A has the following form:*

$$\mathrm{A} = \begin{pmatrix} \bar{\mathrm{A}} & 0^{m-1} \\ \mathrm{w} & 1 \end{pmatrix},$$

where 0^{m-1} is an all-zero column of length $m-1$, $\bar{\mathrm{A}}$ is an arbitrary nonsingular $(m-1) \times (m-1)$ matrix and w is an arbitrary row of length $m-1$.

Proof. $\Longleftarrow$

Trivially, such transformation of the first $m-1$ variables keeps $\overline{x_m}$ in f the only monomial of degree $(m-1)$, and linear transformation cannot increase the degree of h.

$\Longrightarrow$

Assume that $f \circ L_{\mathrm{A}}$ is of the form $\overline{x_m} + h_1$ for some h_1 of degree at most $m-2$, as described in the lemma. This means that the change of variables keeps the monomial $\overline{x_m}$ intact and does not produce any other monomials of degree $m-1$. Clearly, the action of this change on the function h is irrelevant, so let us look at the action on $\overline{x_m}$. It is easy to see that the coefficient of the monomial $\overline{x_i}$ in the resulting function, obtained after applying transformation A to the variables, is precisely the value of a $(m-1) \times (m-1)$ minor, obtained from the matrix A by removing m-th row and i-th column. So we need matrix A to have all such minors be equal to zero, except for the last one, obtained by removing the last column.

Let us denote as $\bar{\mathrm{A}}^1, \bar{\mathrm{A}}^2, \ldots, \bar{\mathrm{A}}^m$ the columns of the matrix A with the last coordinate removed. Denote as $\bar{\mathrm{A}}$ the $(m-1) \times (m-1)$ matrix composed of the first $m-1$ of these columns. Then the condition on the minors can be reformulated as follows: sets of vectors $\{\bar{\mathrm{A}}^1, \ldots, \bar{\mathrm{A}}^{i-1}, \bar{\mathrm{A}}^{i+1}, \ldots, \bar{\mathrm{A}}^m\}$ are linearly dependent for all $i \neq m$, while $\bar{\mathrm{A}}$ is nonsingular. This implies that the following set of equations holds:

$$\begin{cases} \bar{\mathrm{A}}^m + \sum\limits_{j \leqslant m-1} b_{1,j} \bar{\mathrm{A}}^j = 0 \\ \bar{\mathrm{A}}^m + \sum\limits_{j \leqslant m-1} b_{2,j} \bar{\mathrm{A}}^j = 0 \\ \ldots \\ \bar{\mathrm{A}}^m + \sum\limits_{j \leqslant m-1} b_{m-1,j} \bar{\mathrm{A}}^j = 0 \end{cases}$$

where $\mathrm{B} = (b_{i,j})$ is some $(m-1) \times (m-1)$ matrix with $b_{i,i} = 0$ for all i. If we denote rows of the matrix B as B_i, then we can rewrite this in the following manner:

$$\begin{cases} \bar{\mathrm{A}} \cdot \mathrm{B}_1^T = \bar{\mathrm{A}}^m \\ \bar{\mathrm{A}} \cdot \mathrm{B}_2^T = \bar{\mathrm{A}}^m \\ \ldots \\ \bar{\mathrm{A}} \cdot \mathrm{B}_{m-1}^T = \bar{\mathrm{A}}^m \end{cases}$$

Since $\bar{\mathrm{A}}$ is nonsingular, the solution to each equation (which is a system of equations on $b_{i,j}$'s for i-th row) is unique and hence $\mathrm{B}_1 = \mathrm{B}_2 = \ldots = \mathrm{B}_{m-1}$. Since $b_{i,i} = 0$, the matrix B is a zero matrix, which leads to $\bar{\mathrm{A}}^m = 0$. This implies that the last column of A can have 1 only in the last coordinate, and because A is nonsingular, this must be the case. Thus, A is of the form stated in the lemma. □

Connections between quaternary and Boolean bent functions

Natalia Tokareva · Alexander Shaporenko · Patrick Solé

Abstract Boolean bent functions were introduced by Rothaus (1976) as combinatorial objects related to difference sets, and have since enjoyed a great popularity in symmetric cryptography and low correlation sequence design. In this paper connections between classical Boolean bent functions, generalized Boolean bent functions (Schmidt, 2007) and quaternary bent functions (Kumar, Scholtz, Welch, 1985) are studied. We also study Gray images of bent functions and notions of generalized nonlinearity for functions that are relevant to generalized linear cryptanalysis.

Keywords Boolean functions · generalized Boolean functions · quaternary functions · bent functions · semi bent functions · nonlinearity · linear cryptanalysis · Gray map · $\mathbb{Z}_4$-linear codes

Mathematics Subject Classification (2000) 06E30 · 11T71 · 14G50

1 Introduction

Boolean bent functions were introduced by Rothaus [25] as combinatorial objects related to difference sets, and have since enjoyed a great popularity in symmetric cryptography and sequence design. They are, in particular, maps from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$ with some special spectral properties. Their importance in symmetric cryptography stems from linear cryptanalysis of stream ciphers [17–19]. In that context bent functions are the ones which are the worst approximated by affine functions, or, equivalently have the best possible nonlinearity. More information concerning bent functions can be found in monograph of Mesnager [21]. Several researchers [3, 7, 22, 23] have explored extensions

Natalia Tokareva
Sobolev Institute of Mathematics, 4 Acad. Koptyug avenue, 630090 Novosibirsk, Russia
E-mail: tokareva@math.nsc.ru

Alexander Shaporenko
Novosibirsk State University, 2 Pirogova street, 630090 Novosibirsk, Russia
E-mail: shaporenko.alexandr@gmail.com

Patrick Solé
I2M, CNRS, Aix-Marseille University, Centrale Marseille, Marseilles, France
E-mail: patrick.sole@telecom-paristech.fr

of linear cryptanalysis to groups other than the usual elementary abelian 2-groups. In this paper we study a notion of nonlinearity that seems consistent with their notions. We discuss the connection between two notions of $\mathbb{Z}_4$-bentness introduced from a sequence design viewpoint (for applications in CDMA systems) and the classical notion of bent function.

The first approach is to consider functions from $\mathbb{Z}_q^n$ to $\mathbb{Z}_q$, q is any integer, see the paper [11] of Kumar, Scholtz and Welch. We call them **q-ary functions**. Another, more recent approach, which is more natural from the viewpoint of cyclic codes over rings is to consider functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_q$. This is the approach of Schmidt in [26]. We call these latter functions **generalized Boolean functions**. In this paper we focus on the quaternary case ($q = 4$), and explore the interplay between the three types of definitions for bentness.

Let us note that there exist other ways to generalize the concept of bent function. For example, to study bent functions on a finite abelian group [13,31] (later these results were rediscovered in [4]), etc. See a survey of distinct generalizations in [34] and [35].

The material is organized as follows. Necessary definitions are given in section 2. In section 3 we prove that a generalized Boolean function $f(x, y) = a(x, y) + 2b(x, y)$ is bent if and only if Boolean functions b and $a \oplus b$ are both bent. Section 4 shows that there is no direct link between notions of Boolean and quaternary bent functions but we obtain several facts related to bent Boolean and quaternary functions. There is no direct connection between notions of quaternary and generalized bent functions either, which is shown in section 5. Then in section 6 we show that quaternary generalized Boolean bent functions in n variables yield Boolean bent functions by Gray map, or semi bent functions, depending on the parity of n. Section 7 characterizes bent functions by their nonlinearity. Section 8.1 illustrates our results by a survey of the known constructions of generalized bent functions and their Gray images. In section 8.2 we introduce two simple constructions for quaternary bent functions.

Note that a variant of this paper appeared at ePrint archive [29] already in 2009 (see also [30]) but it included an incorrect result about the connection between quaternary and Boolean bent functions (see Lemma 33 and Theorem 34 in [29]). That is why that variant of the paper was unpublished till now although it aroused an interest between specialists. In this paper we correct that mistake and offer new results related to quaternary and Boolean bent functions. After appearence [29] at ePrint archive several related results were obtained by different authors. Thus, Stănică et al. [32] extended the results of [29] related to generalized Boolean bent functions by considering functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_8$. Later the results were extended for functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_{16}$ by Martinsen et al. [15]. Finally, Hodžić et al. [9] gave a complete characterization of generalized bent functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_{2^k}$ for $k > 1$ in terms of both the necessary and sufficient conditions their component Boolean functions need to satisfy. Two open problems that were mentioned in the original paper [29] were solved. More specifically, in [32] the quaternary analogue of Dillon's construction was presented. Then Li et all. [12] characterized the functions in n variables of the form $f(x) = Tr(ax + 2bx^{1+2^k})$ for odd $n/gcd(n/k)$. The results obtained in the original paper [29] were instrumental in the following works [5,6,12,20,24]. The original paper [29] was also mentioned in [16, 28,33]. It remains to note that the present paper contains new results on connections between Boolean, generalized Boolean and quaternary bent functions that were not presented in [29].

2 Definitions and Notation

In what follows by $\oplus$ we mean addition over $\mathbb{Z}_2$ (modulo 2). We will use $+$ for two types of addition: over $\mathbb{Z}_4$ and natural one. It always depends on the context.
We will also use the following two types of inner product:

$$\langle x, y\rangle = x_1y_1 \oplus ... \oplus x_ny_n,$$

$$x.y = x_1y_1 + ... + x_ny_n.$$

Let n, q be integers, $q \geqslant 2$.

We consider the following mappings:

1) $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ — **Boolean function** in n variables. Its *sign function* is $F := (-1)^f$. The *Walsh Hadamard transform* (WHT) of f is

$$\widehat{F}(x) := \sum_{y\in\mathbb{Z}_2^n} (-1)^{f(y)\oplus\langle x,y\rangle} = \sum_{y\in\mathbb{Z}_2^n} F_y(-1)^{\langle x,y\rangle}. \quad (1)$$

A Boolean function f is said to be *bent*, iff $|\widehat{F}(x)| = 2^{n/2}$ for all $x \in \mathbb{Z}_2^n$. It is *semi bent* iff $\widehat{F}(x) \in \{0, \pm 2^{(n+1)/2}\}$ (sometimes such functions are called *near bent*). This is a partial case of *plateaued functions* [36]. Note that Boolean bent (resp. semi bent) functions exist only if the number of variables, n, is even (resp. odd).

2) $f : \mathbb{Z}_2^n \to \mathbb{Z}_q$ — **generalized Boolean function** in n variables. Its *sign function* is $F := \omega^f$, with ω a primitive complex root of unity of order q, i. e. $\omega = e^{2\pi i/q}$. When $q = 4$, we write $\omega = i$. Its WHT is given as

$$\widehat{F}(x) := \sum_{y\in\mathbb{Z}_2^n} \omega^{f(y)}(-1)^{\langle x,y\rangle} = \sum_{y\in\mathbb{Z}_2^n} F_y(-1)^{\langle x,y\rangle}. \quad (2)$$

As above, a generalized Boolean function f is *bent*, iff $|\widehat{F}(x)| = 2^{n/2}$ for all $x \in \mathbb{Z}_2^n$. In comparison to the previous case it does not follow that n should be even if f is bent. Such functions for $q = 4$ were studied by K.-U. Schmidt (2006) in his paper [26]. Here we consider only this partial case $q = 4$.

3) $f : \mathbb{Z}_q^n \to \mathbb{Z}_q$ — **q-ary function** in n variables. Its *sign function* is given by $F := \omega^f$ as in the previous case. Its WHT is defined by

$$\widehat{F}(x) := \sum_{y\in\mathbb{Z}_q^n} \omega^{f(y)+x.y} = \sum_{y\in\mathbb{Z}_q^n} F_y\omega^{x.y}. \quad (3)$$

Here $+$ and $x.y$ are addition and inner product over $\mathbb{Z}_q$. Note that the matrix of this transform is no longer a Sylvester type Hadamard matrix as in the previous case, but a generalized (complex) Hadamard matrix. A q-ary function f is called *bent*, iff $|\widehat{F}(x)| = q^{n/2}$ for all $x \in \mathbb{Z}_q^n$. Notice that again it does not follow from the definition that q-ary bent functions do not exist if n is odd. P. V. Kumar, R. A. Scholtz and L. R. Welch [11] have studied q-ary bent functions in 1985. They proved that such functions exist for any even n and $q \neq 2 (\mathrm{mod}\, 4)$. Later S. V. Agievich [1] proposed an approach to describe regular q-ary bent functions in terms of bent rectangles. If $q = 4$ we call f a **quaternary function**. Here we study such functions only. Note that in 1994 A. S. Ambrosimov [2] studied another type of q-ary bent functions defined over the finite field.

A bent function $f : \mathbb{Z}_q^n \to \mathbb{Z}_q$ is called **regular** if each of its Walsh Hadamard coefficients can be expressed as

$$\widehat{F}(z) = q^{n/2}\omega^{h(z)}$$

for every $z \in \mathbb{Z}_q^n$ and some q-ary function h. From [11] it is known that for quaternary ($q = 4$) case all bent functions are regular.

3 Connections between Boolean and generalized Boolean bent functions

Let $f : \mathbb{Z}_2^{2n} \to \mathbb{Z}_4$ be any generalized Boolean function. Represent it as $f(x, y) = a(x, y) + 2b(x, y)$, for any $x, y \in \mathbb{Z}_2^n$ where $a, b : \mathbb{Z}_2^{2n} \to \mathbb{Z}_2$ are Boolean functions. In this section we study connection between properties of bentness of generalized Boolean and Boolean functions.

Here and further by $\widehat{A \cdot B}$ we mean WHT of $a \oplus b$. It is natural, since $A \cdot B = (-1)^{a \oplus b}$. In this section and in what follows by $x.y$ we mean inner product over $\mathbb{Z}_4$:

$$x.y = x_1y_1 + ... + x_ny_n \bmod 4.$$

Lemma 31 *Between Walsh Hadamard transforms of f, $a \oplus b$, b there is the relation*

$$|\widehat{F}(x, y)|^2 = \frac{1}{2}\left(\widehat{B}^2(x, y) + \widehat{A \cdot B}^2(x, y)\right).$$

Proof Study the Walsh Hadamard Transform of f. According to (2) we have

$$\widehat{F}(x, y) = \sum_{x',y'} (-1)^{\langle x.x' \rangle \oplus \langle y.y' \rangle \oplus b(x',y')}\, i^{\,a(x',y')}.$$

Applying the formula $i^s = \frac{1+(-1)^s}{2} + \frac{1-(-1)^s}{2} i$ for $s = a(x', y')$ we get

$$\widehat{F}(x, y) = \frac{1}{2}\left(\widehat{B}(x, y) + \widehat{A \cdot B}(x, y)\right) + \frac{i}{2}\left(\widehat{B}(x, y) - \widehat{A \cdot B}(x, y)\right).$$

From this we directly get what we need. □

Note that Lemma 31 holds for any (not only even) number of variables of the function f.

Theorem 32 *The following statements are equivalent*:
(i) *the generalized Boolean function f is bent in $2n$ variables*;
(ii) *the Boolean functions of $2n$ variables b and $a \oplus b$ are both bent.*

Proof By Lemma 31 we have $|\widehat{F}(x, y)|^2 = \frac{1}{2}\left(\widehat{B}^2(x, y) + \widehat{A \cdot B}^2(x, y)\right)$. If $a \oplus b$ and b are bent functions then $|\widehat{F}(x, y)|^2 = \frac{1}{2}(2^{2n} + 2^{2n}) = 2^{2n}$ and f is a bent function. Conversely, if f is bent, then it holds $\widehat{B}^2(x, y) + \widehat{A \cdot B}^2(x, y) = 2^{2n+1}$. Since WHT coefficients of a Boolean function are integer, this equality has the unique solution $\widehat{B}^2(x, y) = \widehat{A \cdot B}^2(x, y) = 2^{2n}$ (see [10] for detail). So, functions $a \oplus b$ and b are bent. □

Note that there are some intersections between Lemma 31, the part (i)→(ii) of Theorem 32 and results of the last version of [26].

4 Connections between Boolean and quaternary bent functions

Define a quaternary function $g : \mathbb{Z}_4^n \to \mathbb{Z}_4$ as $g(x+2y) = a(x,y)+2b(x,y)$, for any $x, y \in \mathbb{Z}_2^n$ where $a, b : \mathbb{Z}_2^{2n} \to \mathbb{Z}_2$ are Boolean functions. In this section we study connection between properties of bentness of quaternary and Boolean functions.

4.1 Preliminaries and necessary statements

In this section we present several facts that will be instrumental in what follows.

Lemma 41 *Let $x, y \in \mathbb{Z}_2^n$. If $x.y \neq \langle x, y\rangle$ then $x.y = \langle x, y\rangle + 2$.*

Proof There are four possible values for $x.y : 0, 1, 2$ and 3. For $x.y = 0$ or 1 it is obvious that $x.y = \langle x, y\rangle$. For two remaning cases we have

$$x.y = 2 \to \langle x, y\rangle = 0 \to x.y = \langle x, y\rangle + 2,$$

$$x.y = 3 \to \langle x, y\rangle = 1 \to x.y = \langle x, y\rangle + 2.$$

□

The following fact is well known for Boolean functions.

Lemma 42 *Let f be a linear Boolean function in n variables. Then there are two possible values of Walsh Hadamard coefficients of f: 0 and 2^n.*

Proof Any linear Boolean function f in n variables can be represented as $f(x) = \langle a, x\rangle$ for $a \in \mathbb{Z}_2^n$. Therefore, by (1)

$$\widehat{F}(x) = \sum_{y \in \mathbb{Z}_2^n} (-1)^{\langle a,y\rangle \oplus \langle x,y\rangle} = \sum_{y \in \mathbb{Z}_2^n} (-1)^{\langle a \oplus x, y\rangle}.$$

Using the well-known fact that

$$\sum_{b \in \mathbb{Z}_2^n} (-1)^{\langle b,c\rangle} = \begin{cases} 2^n, \, if \; c = 0, \\ 0, \quad otherwise. \end{cases}$$

the result follows. □

Proposition 43 *(see, for instance, [35]) All quadratic Boolean functions in two variables, i.e. $f : \mathbb{Z}_2^2 \to \mathbb{Z}_2$ such that $f(x, y) = xy \oplus c$, where $x, y, c \in \mathbb{Z}_2$, are bent.*

Proposition 44 *(Rothaus, [25]) Let $n \geqslant 4$ and even. Then degree of a Boolean bent function f in n variables is not more than $n/2$.*

Proposition 45 *(Rothaus, [25]) Let $x \in \mathbb{Z}_2^r$ and $y \in \mathbb{Z}_2^k$, where $r, k \geqslant 2$ and even. A Boolean function $f(x, y) = f_1(x) \oplus f_2(y)$ is a bent function in $r + k$ variables if and only if the functions f_1 and f_2 are bent functions in r and k variables respectively.*

Proposition 46 *(Singh et al, [27]) Let $x \in \mathbb{Z}_4^r$ and $y \in \mathbb{Z}_4^k$ for $r, k \geqslant 1$. A quaternary function $g(x, y) = g_1(x) \oplus g_2(y)$ is a bent function in $r + k$ variables if and only if functions g_1 and g_2 are quaternary bent functions in r and k variables.*

Note that results of Propositions 45 and 46 can be easily extended to sums with more than two functions.

		Number of quaternary bent functions	
Cases for b and $a \oplus b$	Types of a in the case	For each type of a	Total in the case
b and $a \oplus b$ are nonlinear functions	a is a bent function	49152	147456
	a is a linear function	3072	
	a is a nonliear function	95232	
b and $a \oplus b$ are bent functions	a is a bent function	16384	53248
	a is a linear	2304	
	a is a constant	768	
	a is a nonlinear function	33792	

Table 1: Classification of functions b and $a \oplus b$ for quaternary bent functions in 2 variables.

4.2 Quaternary bent functions in small number of variables

Here we present results on connections between notions of quaternary bent functions in one and two variables and Boolean bent functions. Using computer search we obtain the following facts.

Lemma 47 *For every quaternary function $g(x+2y) = a(x,y)+2b(x,y)$ in one variable with $x, y \in \mathbb{Z}_2$ it is true that g is a quaternary bent function if and only if b is a bent function and a does not depend on y, i.e. $a(x,y) = 0$, 1, x or $x \oplus 1$. Moreover, if g is a bent function then b and $a \oplus b$ are bent functions too.*

Computer search shows that the number of quaternary bent functions in one variable is equal to 32.

There are 200704 quaternary bent functions in 2 variables. Among them there are 98304 fuctions such that none of Boolean functions a, b and $a \oplus b$ is a bent function but for 3072 of them a is a linear Boolean function. There are 36864 quaternary bent functions such that b and $a \oplus b$ are bent functions, while for 33792 of them a is a nonlinear function, and for 2304 and 768 functions a is a linear function or constant respectively. The number of quaternary bent functions in 2 variables with each of a, b and $a \oplus b$ being a bent function is equal to 16384. For the remaining 49152 quaternary functions, a is a bent function and b and $a \oplus b$ are nonlinear Boolean functions.
We summarize the data described above in Table 1.

For functions in three and more variables an exhaustive search is not feasible (there are 2^{128} quaternary functions in three variables).

4.3 Possibilities for bentness

From Lemma 47 we know that if g is quaternary bent then b and $a \oplus b$ are bent functions too. In the previous section we showed that it does not hold for quaternary functions in 2 variables. Let us prove that it does not hold for arbitrary $n \geqslant 2$.

Proposition 48 *For every $n \geqslant 2$ there exists a quaternary bent function $g(x+2y) = a(x,y) + 2b(x,y)$ in n variables, with b and $a \oplus b$ being not bent in $2n$ variables.*

Proof In what follows, '+' denotes the addition over $\mathbb{Z}_4$ excepting summation of indices.

Any quaternary function g in n variables can be uniquely represented as follows

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = a(x_1, ..., x_{2n}) + 2b(x_1, ..., x_{2n}).$$

Let

$$b(x_1, .., x_{2n}) = \bigoplus_{i=3}^{n} x_i x_{i+n} \oplus x_1 x_{n+2} \oplus x_2 x_{n+1} \oplus x_1 x_2 x_{n+1},$$

$$a(x_1, .., x_{2n}) = x_1 x_{n+1}.$$

One can see that b can be divided into sum of $n-2$ Boolean functions in two variables and one Boolean function in four variables

$$b(x_1, ..., x_{2n}) = b_1(x_1, x_2, x_{n+1}, x_{n+2}) \oplus b_2(x_3, x_{n+3}) \oplus ... \oplus b_{n-1}(x_n, x_{2n}),$$

$$b_1(x_1, x_2, x_{n+1}, x_{n+2}) = x_1 x_{n+2} \oplus x_2 x_{n+1} \oplus x_1 x_2 x_{n+1},$$

$$b_i(x_{i+1}, x_{n+i+1}) = x_{i+1} x_{n+i+1}, \quad i = 2, ..., n-1.$$

From Proposition 45 we know that b is bent if and only if all b_i are bent. According to Proposition 44 we get that function b_1 in four variables is not bent because its degree is equal to three. Therefore, b is not bent.

It is easy to check that it holds

$$2b(x_1, .., x_{2n}) = (2x_3 x_{n+3} + ... + 2x_n x_{2n}) + 2x_1 x_{n+2} + 2x_2 x_{n+1} + 2x_1 x_2 x_{n+1}.$$

Moreover, g can be divided into sum of $n-2$ quaternary functions in one variable and one quaternary function in two variables

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = g_1(x_1 + 2x_{n+1}, x_2 + 2x_{n+2}) +$$

$$+ g_2(x_3 + 2x_{n+3}) + ... + g_{n-1}(x_n + 2x_{2n}),$$

where

$$g_1(x_1 + 2x_{n+1}, x_2 + 2x_{n+2}) = x_1 x_{n+1} + 2x_1 x_{n+2} + 2x_2 x_{n+1} + 2x_1 x_2 x_{n+1},$$

$$g_i(x_{i+1} + 2x_{n+i+1}) = 2x_{i+1} x_{n+i+1}, \quad i = 2, ..., n-1.$$

From Proposition 43 we know that all x_{i+1}, x_{n+i+1} are bent, $i = 2, ..., n$. Therefore, according to Lemma 47 functions g_i are quaternary bent functions, $i = 2, ..., n-1$. It was checked that the quaternary function g_1 is also bent according to the definition: its WHT coefficients are the following:

$x \in \mathbb{Z}_4^2$	00	01	02	03	10	11	12	13	20	21	22	23	30	31	32	33
$\widehat{G_1(x)}$	4	$4i$	4	4	4	$4i$	-4	4	4	$-4i$	4	-4	4	$-4i$	-4	-4

From Proposition 46 we know that g is a quaternary bent function if and only if all g_i are quaternary bent functions, $i = 1, ..., n-1$. This completes the proof. □

The next result shows that the bentness of a quaternary function does not follow from bentness of Boolean functions in general.

Proposition 49 *For every $n \geqslant 1$ there exists a quaternary function $g(x + 2y) = a(x, y) + 2b(x, y)$ in n variables that is not bent, while b and $a \oplus b$ are Boolean bent functions in $2n$ variables.*

Proof Any quaternary function g in n variables can be uniquely represented as follows

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = a(x_1, ..., x_{2n}) + 2b(x_1, ..., x_{2n}).$$

Let

$$b(x_1, .., x_{2n}) = \bigoplus_{i=1}^{n} x_i x_{i+n},$$
$$a(x_1, .., x_{2n}) = x_{n+1}.$$

It is easy to check that $2b(x_1, .., x_{2n}) = 2x_1x_{n+1} + ... + 2x_nx_{2n}$. One can see that g can be divided into sum of n quaternary functions in one variable

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = g_1(x_1 + 2x_{n+1}) + ... + g_n(x_n + 2x_{2n}),$$

where

$$g_i(x_i + 2x_{n+i}) = a_i(x_i, x_{n+i}) + 2b_i(x_i, x_{n+i}),\ i = 1, .., n,$$
$$b_i(x_i, x_{n+i}) = x_i x_{n+i},\ i = 1, .., n,$$
$$a_1(x_1, x_{n+1}) = x_{n+1},$$
$$a_i(x_i, x_{n+i}) = 0,\ i = 2, .., n.$$

From Proposition 46 we know that g is a quaternary bent function if and only if all g_i are quaternary bent functions, $i = 1, ..., n$. From Lemma 47 and by the choice of a and b we get that g_1 is not quaternary bent. This completes the proof. □

From Propositions 48 and 49 we conclude that there is no direct link between notions of Boolean and quaternary bent functions. Additionally, Proposition 48 shows that if b and $a \oplus b$ are not bent it does not imply that g is not bent. According to Proposition 49 it is also true that if g is not bent it does not imply that b and $a \oplus b$ are not bent.

From the previous section we can see that for quaternary bent functions in one and two variables a Boolean function b is bent if and only if $a \oplus b$ is also bent. Whether this statement is true for arbitrary n remains an open problem.

4.4 Nonlinearity of component Boolean functions

Let $g(x+2y) = a(x, y)+2b(x, y)$ be a quaternary function in n variabes, where $x, y \in \mathbb{Z}_2^n$ and a, b are Boolean functions in $2n$ variables.

Let us represent WHT coefficients of quaternary functions in terms of the coefficients of Boolean functions b and $a \oplus b$ as we did for generalized functions in section 4. Here by $\widehat{A \cdot B}$ we mean the WHT of $a \oplus b$.

Lemma 410 *Between the WHT coefficients of g, $a \oplus b$, b there is the relation*

$$\widehat{G}(x + 2y) = \frac{1}{2}\left(\widehat{B}(x \oplus y, x) + \widehat{A \cdot B}(y, x) - 2c_{\,b}(x \oplus y, x) - 2c_{\,a\oplus b}(y, x)\right) +$$
$$+\frac{i}{2}\left(\widehat{B}(y, x) - \widehat{A \cdot B}(x \oplus y, x) - 2c_{\,b}(y, x) + 2c_{\,a\oplus b}(x \oplus y, x)\right),$$

with

$$c_{\,f}(u, x) = \sum_{x' \in V_x, y'} (-1)^{f(x', y') \oplus \langle (u,x), (x', y') \rangle},$$

where f is a Boolean function in $2n$ variables, $u \in \mathbb{Z}_2^n$, and $V_x = \{\ x' \mid \langle x, x' \rangle \neq x.x'\ \}$.

Proof Study the Walsh Hadamard Transform of g. By (3) we know that

$$\widehat{G}(x+2y) = \sum_{x',y'} i^{\,(x+2y).(x'+2y')+a(x',y')+2b(x',y')}.$$

From Lemma 41 and the fact that for any $x'', x''' \in \mathbb{Z}_2^n$ it holds $2\langle x'', x'''\rangle = 2x''.x'''$ we have

$$(x+2y).(x'+2y') = \begin{cases} \langle x,x'\rangle + 2\langle x,y'\rangle + 2\langle y,x'\rangle, & if\ x.x' = \langle x,x'\rangle, \\ \langle x,x'\rangle + 2\langle x,y'\rangle + 2\langle y,x'\rangle + 2, & if\ x.x' \neq \langle x,x'\rangle. \end{cases}$$

Let $U_x = \{\ x' \in \mathbb{Z}_2^n \,|\, x.x' = \langle x,x'\rangle\ \}$ and $V_x = \{\ x' \in \mathbb{Z}_2^n \,|\, x.x' \neq \langle x,x'\rangle\ \}$. Therefore, we get $U_x \cap V_x = \varnothing$ and $U_x \cup V_x = \mathbb{Z}_2^n$. Note that $|U_x| \neq |V_x|$ in general. Then

$$\widehat{G}(x+2y) = \sum_{x\in U_x,y'} (-1)^{\langle x,y'\rangle \oplus \langle y,x'\rangle \oplus b(x',y')}\, i^{\,\langle x,x'\rangle + a(x',y')} -$$

$$- \sum_{x'\in V_x,y'} (-1)^{\langle x,y'\rangle \oplus \langle y,x'\rangle \oplus b(x',y')}\, i^{\,\langle x,x'\rangle + a(x',y')}.$$

Here we use the standard maps $\beta, \gamma : \mathbb{Z}_4 \to \mathbb{Z}_2$ defined as

$$\beta : 0,1 \to 0 \text{ and } \beta : 2,3 \to 1;$$

$$\gamma : 0,2 \to 0 \text{ and } \gamma : 1,3 \to 1.$$

For any $t \in \mathbb{Z}_4$ it holds

$$i^t = (-1)^{\beta(t)} \left(\frac{1+(-1)^{\gamma(t)}}{2} + \frac{1-(-1)^{\gamma(t)}}{2} i \right).$$

Using this formula for $t = x.x' + a(x',y')$ and the fact that $\gamma(\langle x,x'\rangle + a(x',y')) = \langle x,x'\rangle \oplus a(x',y')$ we get

$$\widehat{G}(x+2y) = \frac{1}{2}\left(S_1 + S_2 - S_3 - S_4\right) + \frac{i}{2}\left(S_1 - S_2 - S_3 + S_4\right),$$

where

$$S_1 = \sum_{x'\in U_x,y'} (-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\beta(\langle x,x'\rangle + a(x',y'))},$$

$$S_2 = \sum_{x'\in U_x,y'} (-1)^{a(x',y')\oplus b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle\oplus\beta(\langle x,x'\rangle + a(x',y'))},$$

$$S_3 = \sum_{x'\in V_x,y'} (-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\beta(\langle x,x'\rangle + a(x',y'))},$$

$$S_4 = \sum_{x'\in V_x,y'} (-1)^{a(x',y')\oplus b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle\oplus\beta(\langle x,x'\rangle + a(x',y'))}.$$

Let $M_{\delta,x} = \{\ x' \in \mathbb{Z}_2^n \,|\, \langle x,x'\rangle = \delta\ \}$ for $\delta \in \mathbb{Z}_2$. Note that $M_{0,x} \cup M_{1,x} = \mathbb{Z}_2^n$ and $|M_{0,x}| = |M_{1,x}| = 2^{n-1}$. Let us divide every sum S_1, S_2, S_3 and S_4 into two sums

$\sum_{x'\in M_{0,x},y'}$ and $\sum_{x'\in M_{1,x},y'}$. Note that $\beta(a(x',y')+\langle x,x'\rangle)$ is equal to 0 or $a(x',y')$ for $x'\in M_{0,x}$ and $x'\in M_{1,x}$ respectively. Thus, we have

$$S_1=\sum_{x'\in U_x\cap M_{0,x},y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle}+$$
$$+\sum_{x'\in U_x\cap M_{1,x},y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus a(x',y')},$$
$$S_2=\sum_{x'\in U_x\cap M_{0,x},y'}(-1)^{a(x',y')\oplus b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle}+$$
$$+\sum_{x'\in U_x\cap M_{1,x},y'}(-1)^{a(x',y')\oplus b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle\oplus a(x',y')},$$
$$S_3=\sum_{x'\in V_x\cap M_{0,x},y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle}+$$
$$+\sum_{x'\in V_x\cap M_{1,x},y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus a(x',y')},$$
$$S_4=\sum_{x'\in V_x\cap M_{0,x},y'}(-1)^{a(x',y')\oplus b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle}+$$
$$+\sum_{x'\in V_x\cap M_{1,x},y'}(-1)^{a(x',y')\oplus b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle a(x',y')}.$$

After grouping terms we obtain

$$S_1+S_2-S_3-S_4=$$
$$=\sum_{x'\in U_x,y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle}+$$
$$+\sum_{x'\in U_x,y'}(-1)^{b(x',y')\oplus a(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle}-$$
$$-\sum_{x'\in V_x,y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle}-$$
$$-\sum_{x'\in V_x,y'}(-1)^{b(x',y')\oplus a(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle}.$$

Then

$$S_1-S_2-S_3+S_4=$$
$$=\sum_{x'\in U_x,y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle}-$$
$$-\sum_{x'\in U_x,y'}(-1)^{b(x',y')\oplus a(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle}-$$
$$-\sum_{x'\in V_x,y'}(-1)^{b(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle}+$$

$$+\sum_{x'\in V_x,y'}(-1)^{b(x',y')\oplus a(x',y')\oplus\langle x,y'\rangle\oplus\langle y,x'\rangle\oplus\langle x,x'\rangle}.$$

Since

$$c_f(u,x)=\sum_{x'\in V_x,y'}(-1)^{f(x',y')\oplus\langle(u,x),(x',y')\rangle},$$

where f is a Boolean function in $2n$ variables and $u\in\mathbb{Z}_2^n$, then one can see that

$$S_1+S_2-S_3-S_4=$$

$$=(\widehat{B}(x\oplus y,x)-c_b(x\oplus y,x))+(\widehat{A\cdot B}(y,x)-c_{a\oplus b}(y,x))-c_b(x\oplus y,x)-c_{a\oplus b}(y,x)$$

and

$$S_1-S_2-S_3+S_4=$$

$$=(\widehat{B}(y,x)-c_b(y,x))-(\widehat{A\cdot B}(x\oplus y,x)-c_{a\oplus b}(x\oplus y,x))-c_b(y,x)+c_{a\oplus b}(x\oplus y,x).$$

After rearranging, the result follows. □

We can see that WHT coefficients of a quaternary function g do not directly depend on WHT coefficients of Boolean functions b and $a\oplus b$. This result will not be useful in studying connection between bentness of quaternary and Boolean functions but it will be instrumental for the next result and also in section 8.2.

Theorem 411 *Let $g(x+2y)=a(x,y)+2b(x,y)$ be a quaternary bent function with $x,y\in\mathbb{Z}_2^n$ and a,b be Boolean functions in $2n$ variables. Then b and $a\oplus b$ are nonlinear functions for any $n\geqslant 1$.*

Proof According to Lemma 42 there are two possible values of Walsh Hadamard coefficients of a linear Boolean function in $2n$ variables: 0 and 2^{2n}.

From Lemma 410 we get that

$$\widehat{G}(2y)=\frac{1}{2}(\widehat{B}(y,0)+\widehat{A\cdot B}(y,0))+\frac{i}{2}(\widehat{B}(y,0)-\widehat{A\cdot B}(y,0)),\text{ where } y\in\mathbb{Z}_2^n.$$

Note that the reason why there are no coefficients $c_b(x\oplus y,x), c_b(y,x), c_{a\oplus b}(x\oplus y,x)$ and $c_{a\oplus b}(y,x)$ is because the set V_x is empty for $x=\mathbf{0}$.

As it was mentioned in section 2 all quaternary bent functions are regular. It means that there is only real or imaginary part of $\widehat{G}(2y)$. Thus, we get that there are two possible cases

$$\begin{cases}(\widehat{B}(y,0)+\widehat{A\cdot B}(y,0))^2=0,\\(\widehat{B}(y,0)-\widehat{A\cdot B}(y,0))^2=4\cdot 4^n.\end{cases}$$

or

$$\begin{cases}(\widehat{B}(y,0)+\widehat{A\cdot B}(y,0))^2=4\cdot 4^n,\\(\widehat{B}(y,0)-\widehat{A\cdot B}(y,0))^2=0.\end{cases}$$

From the first system we get

$$\begin{cases}\widehat{B}(y,0)=-\widehat{A\cdot B}(y,0),\\(2\cdot\widehat{B}(y,0))^2=4\cdot\widehat{B}(y,0)^2=4\cdot 4^n.\end{cases}$$

Hence,

$$\widehat{B}(y,0)=-\widehat{A\cdot B}(y,0)=\pm 2^n.$$

By solving the second system one can get

$$\widehat{B}(y,0)=\widehat{A\cdot B}(y,0)=\pm 2^n.$$

Therefore, b and $a\oplus b$ are nonlinear functions. □

5 Connections between quaternary and generalized Boolean bent functions

Let $g(x+2y) = f(x,y)$, where $g : \mathbb{Z}_4^n \to \mathbb{Z}_4$, $f : \mathbb{Z}_2^n \to \mathbb{Z}_4$ and $x, y \in \mathbb{Z}_2^n$.

In this section we will show that approach of Kumar et al and that of Schmidt are not equivalent.

Proposition 51 *For every $n \geqslant 1$ there exists a generalized bent function $f(x,y)$ in $2n$ variables such that a quaternary function $g(x+2y)$ in n variables defined as $g(x+2y) = f(x,y)$ for all $x, y \in \mathbb{Z}_2^n$ is not bent.*

Proof From Proposition 49 there exists a quaternary function $g(x+2y) = a(x,y) + 2b(x,y)$ which is not bent, while b and $a \oplus b$ are both bent.

Now from Theorem 32 we know that if b and $a \oplus b$ are both bent then $f(x,y)$ is a generalized bent function. □

Proposition 52 *For every $n \geqslant 2$ there exists a quaternary bent function $g(x+2y)$ in n variables such that a generalized function $f(x,y)$ in $2n$ variables defined as $f(x,y) = g(x+2y)$ for all $x, y \in \mathbb{Z}_2^n$ is not bent.*

Proof From Proposition 48 for $n \geqslant 1$ there exists a quaternary bent function $g(x+2y) = a(x,y) + 2b(x,y)$ in n variables such that b and $a \oplus b$ are both not bent.

From Theorem 32 we know that generalized function $f(x,y)$ is bent iff b and $a \oplus b$ are both bent. Hence, $f(x,y)$ is not bent. □

6 Gray images of bent functions

Let f be a generalized Boolean function from $\mathbb{Z}_2^n$ to $\mathbb{Z}_4$. Write $f = a + 2b$ with a, b Boolean functions in n variables. Its *Gray map* $\phi(f)$ is the Boolean function in variables (x,z) with $x \in \mathbb{Z}_2^n$ and $z \in \mathbb{Z}_2$ defined as $a(x)z + b(x)$. The proof of the next result is implicit in the proof of [26, Th. 3.5] and is omitted.

Proposition 61 *For the WHTs of functions f and $\phi(f)$ it holds*

$$\widehat{\Phi(f)}(u,v) = 2\Re(i^{-v}\widehat{F}(u)) = \widehat{B}(u) + (-1)^v \widehat{A \cdot B}(u), \text{ where } u \in \mathbb{Z}_2^n, v \in \mathbb{Z}_2. \quad (4)$$

Here $\Re$ denotes real part of a complex number. As far as the left side of equation (4) is a WHT coefficient of a Boolean function, we easily get

Corollary 62 *For any generalized Boolean function f in n variables it holds*

$$\max_{u \in \mathbb{Z}_2^n, v \in \mathbb{Z}_2} |\Re(i^{-v}\widehat{F}(u))| \geqslant 2^{(n-1)/2}.$$

Corollary 63 *If f is bent in n variables then $\phi(f)$ is either bent (n odd) or semi bent (n even).*

Proof Write $\widehat{F}(u) = X + iY$ with X, Y integers. We know that $2^n = X^2 + Y^2$. We know that the solution to that diophantine equation in $X > 0$ and $X \geqslant Y \geqslant 0$ is unique, see e.g. [10]. The obvious solutions for n odd are $\{|X| = |Y| = 2^{(n-1)/2}\}$, $\{Y = 0,\ X = \pm 2^{n/2}\}$ and $\{Y = \pm 2^{n/2},\ X = 0\}$ for n even.

Thus, if n is odd it holds $\widehat{\Phi(f)}(u,v) = \pm 2^{(n+1)/2}$ for all u, v, and hence $\phi(f)$ is bent in $n+1$ variables. If n is even we see that $\widehat{\Phi(f)}(u,v)$ equals 0 or $\pm 2^{(n+2)/2}$, so $\phi(f)$ is semi bent in $n+1$ variables. □

There is a partial converse to Corollary 63. The proof is immediate.

Proposition 64 *Let n be odd. If $\phi(f)$ is a Boolean bent function in $n+1$ variables then f is a generalized Boolean bent function in n variables.*

This fact has also been obtained in the last variant of [26].

7 Notions of nonlinearity

It is well-known that Boolean bent functions are characterized by their maximal distance to the first order Reed Muller code. This fact is generalized in this section to their quaternary analogues.

7.1 Generalized Boolean functions

Let $RM(r,k)$ be the Reed Muller code of length 2^k and of order r, see [14]. Define, for $0 \leqslant r \leqslant m$ the quaternary code $ZRM(r,m) = \phi^{-1}(RM(r,m+1))$. This code is spanned by vectors of values for functions of degree at most $r-1$ together with twice functions of degree at most r, see [8] for detail. We introduce the **nonlinearity** $N(f)$ of a generalized bent Boolean function f in n variables as

$$N(f) := 2^n - \frac{1}{2} \max_{u \in \mathbb{Z}_2^n, v \in \mathbb{Z}_2} |\widehat{\Phi(f)}(u,v)|. \tag{5}$$

We denote by $d_L(\cdot,\cdot)$ the Lee distance on $\mathbb{Z}_4^N$. Analogously, let $d_H(\cdot,\cdot)$ be the Hamming distance on $\mathbb{Z}_2^{2N}$. According to Corollary 62 we have

Proposition 71 *For any generalized Boolean function f in n variables it is true $N(f) \leqslant 2^n - 2^{(n-1)/2}$.*

Proposition 72 *With the above notation, for any generalized Boolean function in n variables f we have*

$$N(f) = d_L(f, ZRM(1,n)) = d_H(\Phi(f), RM(1,n+1)).$$

Proof Let x, y be arbitrary vectors of $\mathbb{Z}_4^N$. Denote by i^x the vector $(i^{x_1}, \ldots, i^{x_N})$. Recall first the well-known identities

$$d_E^2(i^x, i^y) = 2d_L(x,y) = 2(N - \Re(\sum_{j=1}^{N} i^{x_j - y_j})),$$

where d_E stands for the Euclidean distance. Observe that $ZRM(1,n)$ is spanned by the all-one vector, along with twice the binary linear functions, and that $\widehat{F}(u) = \sum_{y \in \mathbb{Z}_2^n} i^{f(y)+2u.y}$. The second equality holds by the isometry property of the Gray map [8]. □

Hence, using Propositions 71 and 72 we can reformulate one partial case from Corollary 63 and Proposition 64 as follows.

Corollary 73 *Let n be odd. A function f is bent if and only if $N(f)$ attains the maximal possible value $2^n - 2^{(n-1)/2}$.*

The case of even n is more complicated. We have

Corollary 74 *Let n be even. If a function f is bent then $N(f) = 2^n - 2^{n/2}$.*

Proof By Corollary 63 the Boolean function $\phi(f)$ is semi bent in $n+1$ variables. Hence the maximum value of $|\widehat{\Phi(f)}(u,v)|$ is equal to $2^{(n+2)/2}$. Then by Proposition 61 and definition (5) we get $N(f) = 2^n - 2^{n/2}$. □

The converse statement is not right in general as far as from the equality

$$\max_{u\in\mathbb{Z}_2^n, v\in\mathbb{Z}_2} |\widehat{\Phi(f)}(u,v)| = 2^{(n+2)/2}$$

it does not follow that $|\widehat{F}(u)| = 2^{n/2}$ for any $u \in \mathbb{Z}_2^n$. Actually, it is not clear what is the maximum possible value of $N(f)$ if n is even. To know it one should find the value of covering radius of the code $RM(1, n+1)$ when $n+1$ is odd. But it is a hard old problem without analogy to the easy case of even $n+1$.

7.2 Quaternary functions

Let g be a quaternary function in n variables. In this case, an immediate reduction to the preceding subsection (namely, passing from g to f in the notations of section 5) yields the definition

$$N(g) := 2^{2n} - \frac{1}{2} \max_{u,v\in\mathbb{Z}_2^n, w\in\mathbb{Z}_2} |\widehat{\Phi(g)}(u,v,w)|.$$

The following analogue of Proposition 72 is immediate.

Proposition 75 *For any quaternary function g in n variables we have*

$$N(g) = d_L(g, ZRM(1,2n)) = d_H(\phi(g), RM(1,2n+1)).$$

In particular if g is bent then $N(g) = 2^{2n} - 2^n$. As it was mentioned above the maximal possible value of $N(g)$ is not determined yet.

8 Examples of Constructions

The degree of a generalized Boolean function f denoted by $\deg(f)$ is understood in the sense of its algebraic normal form (ANF). For computing degrees we require the following lemma.

Lemma 81 *For a generalized Boolean function f the degree of $\phi(f)$ is at most the degree of f.*

Proof Follows by definition of the $ZRM(r,m)$ code by its generators [8]. □

8.1 Generalized Boolean bent functions

In [26, Th. 4.3] figures a natural generalization of the classical Maiorana McFarland construction.

Proposition 82 *(Schmidt, [26]) The generalized Boolean function f in $2n$ variables defined for x, y in $\mathbb{Z}_2^n$ by*

$$f(x,y) = 2x.\pi(y) + \tau(y),$$

with τ an arbitrary generalized Boolean function in n variables and π an arbitrary permutation of $\mathbb{Z}_2^n$ is bent.

By Corollary 63 the Gray map of this function is a binary Boolean semi bent function in $2n+1$ variables. By Lemma 81 its degree is $\max(2, \deg(\tau))$.

It is well-known that the binary Kerdock code contains bent functions. We assume the reader has some familiarity with Galois rings as can be gained in, e.g. [8].

Proposition 83 *(Schmidt, [26]) Let $n \geqslant 3$ denote an integer. Let R_n denote the Galois ring of characteristic 4 and size 4^n. Let R_n^x denote $R_n \setminus 2R_n$. Let T_n denote the Teichmuller set of R_n, and Tr the trace function of R_n. The generalized Boolean function in n variables defined for $x \in T_n$ by*

$$f(x) = \epsilon + Tr(sx)$$

for constants ϵ, s ranging in $\mathbb{Z}_4, R_n^x$ is bent. Its Gray image is either bent (n odd) or semi bent (n even).

Proof The first assertion follows by [26, Construction 5.2] upon observing that $ZRM(1,n)$ is described by functions $f(x) = \epsilon + 2Tr(sx)$. The second assertion follows by Corollary 63. □

A monomial construction of a bent generalized Boolean function is in [26, Th. 5.3]. Intuitively it detects the generalized bent functions in the dual of the Goethals code.

Proposition 84 *(Schmidt, [26]) Keep the notation of Proposition 83. Let μ denote the "reduction mod 2" map from R_n to $\mathbb{F}_{2^n}$. The generalized Boolean function in n variables defined for $x \in T_n$ by*

$$f(x) = \epsilon + Tr(sx + 2tx^3)$$

for constants ϵ, s, t ranging in $\mathbb{Z}_4, R_n, T_n \setminus \{0\}$ is bent if $\mu(s) = 0$ and the equation

$$\mu(t)z^3 + 1 = 0$$

has no solutions in $\mathbb{F}_{2^n}$, or if $\mu(s) \neq 0$ and the equation

$$z^3 + z + \frac{\mu(t)^2}{\mu(t)^6} = 0$$

has no solutions in $\mathbb{F}_{2^n}$.

By Corollary 63 the Gray map of this function is a binary Boolean function in $n+1$ variables which is semi bent if n is even or bent if n is odd. It is quadratic by Lemma 81.

In the original paper [29] it was mentioned that it would be interesting, for instance, to replace the exponent 3 in Proposition 84 by a Gold exponent $2^k + 1$. Then Li et all. [12] characterized the functions in n variables of the form $f(x) = Tr(ax + 2bx^{1+2^k})$ for odd $n/gcd(n/k)$.

Proposition 85 *For every n a quaternary function*

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = cx_j + x_1x_{n+1} + ... + x_nx_{2n}$$

is a quaternary bent function with $c \in \mathbb{Z}_2$, $j \in \{1, ..., n\}$ and '+' is addition over $\mathbb{Z}_4$.

Proof One can see that g can be divided into sum of n quaternary functions in one variable

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = g_1(x_1 + 2x_{1+n}) + ... + g_n(x_n + 2x_{2n}),$$

$$g_i(x_i + 2x_{i+n}) = \begin{cases} 2x_ix_{i+n}, & if\ i \neq j, \\ cx_j + 2x_jx_{j+n}, & if\ i = j. \end{cases}$$

From Proposition 43 we know that all x_i, x_{i+n} are bent, $i = 2, ..., n$. From Lemma 47 each of g_i is a quaternary bent function in one variable, therefore, from Proposition 46 g is also a quaternary bent function. □

Proposition 86 *Let $g(x+2y) = a(x,y)+2b(x,y)$ and $g'(x+2y) = a(x,y)+2(a(x,y)\oplus b(x,y))$ be quaternary functions with $x, y \in \mathbb{Z}_2^n$ and a, b be Boolean functions in $2n$ variables. Then g is bent iff g' is bent.*

Proof Study the Walsh Hadamard Transform of g and g'. From Lemma 410 we have

$$\widehat{G}(x + 2y) = \frac{1}{2}\left(\widehat{B}(x \oplus y, x) + \widehat{A \cdot B}(y, x) - 2c_{\,b}(x \oplus y, x) - 2c_{\,a\oplus b}(y, x)\right) +$$

$$+\frac{i}{2}\left(\widehat{B}(y, x) - \widehat{A \cdot B}(x \oplus y, x) - 2c_{\,b}(y, x) + 2c_{\,a\oplus b}(x \oplus y, x)\right)$$

and

$$\widehat{G'}(x + 2(x \oplus y)) = \frac{1}{2}\left(\widehat{A \cdot B}(y, x) + \widehat{B}(x \oplus y, x) - 2c_{\,b}(y, x) - 2c_{\,a\oplus b}(x \oplus y, x)\right) +$$

$$+\frac{i}{2}\left(\widehat{A \cdot B}(x \oplus y, x) - \widehat{B}(y, x) + 2c_{\,b}(y, x) - 2c_{\,a\oplus b}(x \oplus y, x)\right),$$

with

$$c_{\,f}(u, x) = \sum_{x' \in V_x, y'} (-1)^{f(x',y') \oplus \langle (u,x),(x',y') \rangle},$$

where f is a Boolean function in $2n$ variables, $u \in \mathbb{Z}_2^n$, and $V_x = \{\ x' \mid \langle x, x' \rangle \neq x.x'\ \}$.

Let $\Re$ and $\Im$ be real and imaginary parts of a complex number respectively. Then $\Re(\widehat{G}(x + 2y)) = \Re(\widehat{G'}(x + 2(x \oplus y)))$ and $\Im(\widehat{G}(x + 2y)) = -\Im(\widehat{G'}(x + 2(x \oplus y)))$.

As it was mentioned in section 2 all quaternary bent functions are regular. Therefore, each of Walsh Hadamard coefficients of a quaternary bent function has only real or imaginary part. Hence, if g is bent then $|\widehat{G'}(x + 2(x \oplus y))| = |\widehat{G}(x + 2y)| = 4^{n/2}$. By the same way we can proof that if g' is bent then $|\widehat{G}(x+2y)| = |\widehat{G'}(x+2(x\oplus y))| = 4^{n/2}$. This completes the proof. □

9 Conclusion and open problems

In the present work we have shown how generalizations of the notion of bent functions involving the ring $\mathbb{Z}_4$ could produce, by Gray map or by base 2 expansion, bent Boolean functions in the classical sense. We have proved that the approach of Kumar et al and that of Schmidt are not equivalent at least in quaternary case. Schmidt's definition fits better $\mathbb{Z}_4$-cyclic codes constructions. Conversely classical binary bent functions (but perhaps not semi bent functions) can yield generalized bent functions by inverse Gray map. These results motivate to explore further algebraic constructions of generalized bent functions. Although the results show that there is no direct connection between quaternary and Boolean bent functions it is still might be possible to connect these notions if we will ask for additional conditions. For instance, it would be interesting to solve the problem that we mentioned at the end of section 4.3. It is also possible that notions of $q-ary$ and Boolean bent functions more connected for $q > 4$.

Acknowledgment. Authors wish to thank Sihem Mesnager and Alexander Kutsenko for helpful discussions. The work of the first author was supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research. The second author was supported by RFBR (18-07-01394).

References

1. S. V. Agievich, *Bent rectangles*, NATO Advanced Study Institute on Boolean Functions in Cryptology and Information Security (Zvenigorod, Russia. September 8–18, 2007). Proc: Netherlands, IOS Press. 2008. P. 3–22. Available at `http://arxiv.org/abs/0804.0209`.
2. A. S. Ambrosimov, *Properties of q-ary bent functions over finite fields*, Discrete Mathematics and Applications, 1994, V. 4. N 4. P. 341–350.
3. T. Baignères, P. Junod, S. Vaudenay, *How Far Can We Go Beyond Linear Cryptanalysis?* Advances in Cryptology — ASIACRYPT '04 (Jeju Island, Korea. December 5–9, 2004). Proc. Berlin: Springer, 2004. P. 432–450 (LNCS 3329).
4. C. Carlet, C. Ding, *Highly nonlinear mappings*, J. of Complexity. 2004. V. 20. N 2–3. P. 205–244.
5. S. Gangopadhyay, E. Pasalic and P. Stănică, *A Note on Generalized Bent Criteria for Boolean Functions*, IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 3233–3236, May 2013.
6. S. Gangopadhyay, C. Riera, P. Stănică, *Gowers U2 norm of Boolean functions and their generalizations*, Workshop on Cryptography and Coding, Rennes, France 2019.
7. L. Granboulan, É. Levieil and G. Piret, *Pseudorandom Permutation Families over Abelian Groups*, Fast Software Encryption — FSE 2006 (Graz, Austria. March 15–17, 2006). Springer, 2006. P. 57–77 (LNCS 4047).
8. R. Hammons, V. Kumar, A. R. Calderbank, N. J. A. Sloane, P. Solé, *Kerdock, Preparata, Goethals and others are linear over* $\mathbb{Z}_4$, IEEE Trans. of Information Theory. 1994. V. 40. P. 301–319.
9. S. Hodžić, W. Meidl and E. Pasalic, *Full Characterization of Generalized Bent Functions as (Semi)-Bent Spaces, Their Dual, and the Gray Image*, IEEE Transactions on Information Theory, vol. 64, no. 7, pp. 5432–5440, July 2018.
10. K. Ireland, M. Rosen, *A classical introduction to modern number theory*, GTM 84, Springer. 1990.
11. P. V. Kumar, R. A. Scholtz and L. R. Welch, *Generalized bent functions and their properties*, J. Combin. Theory, v. 40, no. 1, pp. 90–107, 1985.
12. N. Li, X. Tang and T. Helleseth, *New Constructions of Quadratic Bent Functions in Polynomial Form*, IEEE Transactions on Information Theory, vol. 60, no. 9, pp. 5760–5767, Sept. 2014.

13. O. A. Logachev, A. A. Sal'nikov, V. V. Yashenko, *Bent functions on a finite Abelian group*, Discrete Mathematics and Applications. 1997. V. 7. N 6. P. 547–564.
14. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error–Correcting Codes*, Amsterdam: North–Holland, 1977.
15. T. Martinsen, W. Meidl, P. Stănică, *Generalized Bent Functions and Their Gray Images*, Arithmetic of Finite Fields. WAIFI 2016. Lecture Notes in Computer Science. V. 10064. pp. 160–173, 2016.
16. T. Martinsen, W. Meidl, S. Mesnager and P. Stănică, *Decomposing Generalized Bent and Hyperbent Functions*, IEEE Transactions on Information Theory, vol. 63, no. 12, pp. 7804–7812, Dec. 2017.
17. M. Matsui, A. Yamagishi, *A New Method for Known Plaintext Attack of FEAL Cipher*. Advances in Cryptology — EUROCRYPT'92 (Balatonfured, Hungary. May 24–28, 1992). Proc. Berlin: Springer, 1993. P. 81–91 (LNCS 658).
18. M. Matsui, *Linear Cryptanalysis Method for DES Cipher*. Advances in Cryptology — EUROCRYPT'93 (Lofthus, Norway. May 23–27, 1993). Proc. Berlin: Springer, 1994. P. 386–397 (LNCS 765).
19. M. Matsui, *The First Experimental Cryptanalysis of the Data Encryption Standard*. Advances in Cryptology — CRYPTO'94 (Santa Barbara, California, USA. August 21–25, 1994). Proc. Berlin: Springer, 1994. P. 1–11 (LNCS 839).
20. W. Meidl, *A secondary construction of bent functions, octal gbent functions and their duals*, Mathematics and Computers in Simulation, vol. 143, pp. 57–64, Jan. 2018.
21. S. Mesnager, *Bent functions: fundamentals and results*, Springer Verlag, 2016.
22. M. G. Parker and H. Raddum, *Z4-Linear Cryptanalysis*. NESSIE Internal Report, 27/06/2002: NES/DOC/UIB/WP5/018/1.
23. M. G. Parker, *Generalised S-Box Nonlinearity*. NESSIE Public Document, 11.02.03: NES/DOC/UIB/WP5/020/A.
24. C. Riera, P. Stănică, S. Gangopadhyay, *Generalized bent Boolean functions and strongly regular Cayley graphs*, Discrete Applied Mathematics, Feb. 2020.
25. O. Rothaus, *On bent functions*, J. Combin. Theory, Ser. A. 1976. V. 20. N 3. P. 300–305.
26. K-U. Schmidt, *Quaternary Constant-Amplitude Codes for Multicode CDMA*. // IEEE International Symposium on Information Theory — ISIT'2007. (Nice, France. June 24–29, 2007). Proc. 2007. P. 2781–2785. Last version at Trans. Inform. Theory. 2009. V. 55. N 4. P. 1824–1832.
27. D. Singh, M. Bhaintwal, and B. K. Singh, *Some results on q-ary bent functions* // Int. J. Comput. Math., vol. 90, no. 9, pp. 1761–1773, 2013.
28. L. Sok, M. Shi, P. Solé, *Classification and Construction of quaternary self-dual bent functions*, Cryptogr. Commun., 10(2), pp. 277–289 (2018).
29. P. Solé, N. Tokareva, *Connections Between Quaternary and Binary Bent Functions*. // Cryptology ePrint Archive, Report 2009/544, available at `http://eprint.iacr.org/`.
30. P. Solé, N. Tokareva, *On Quaternary and Binary Bent Functions.*, Prikl. Diskr. Mat., 2009, supplement no 1, pp. 16–18. Available at `www.mathnet.ru`.
31. V. I. Solodovnikov *Bent functions from a finite abelian group into a finite abelian group* // Discrete Mathematics and Applications. 2002. V. 12. N 2. P. 111–126.
32. P. Stănică, T. Martinsen, S. Gangopadhyay, B. K. Singh *Bent and generalized bent Boolean functions*// Des. Codes Cryptogr. 2013. V. 69. P. 77–94.
33. C. Tang, C. Xiang, Y. Qi and K. Feng, *Complete Characterization of Generalized Bent and 2k-Bent Boolean Functions*, IEEE Transactions on Information Theory, vol. 63, no. 7, pp. 4668–4674, July 2017.
34. N. N. Tokareva *Generalizations of bent functions. A survey* // Discrete Analysis and Operation Research, 2010. V. 17. N 1. P. 34–64 (in Russian). English translation will appear soon in Journal of Applied and Industrial Mathematics.
35. N. Tokareva, *Bent functions: results and applications to cryptography* // Acad. Press. Elsevier, 2015.
36. Y. Zheng, X.-M. Zhang, *On plateaued functions*, IEEE Trans. of Information Theory, vol. 47. no. 3. pp. 1215–1223, 2001.

МАТЕМАТИЧЕСКИЕ ОСНОВЫ КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ

УДК 004.75

РАЗРАБОТКА МЕТОДА СОКРЫТИЯ ПРИВАТНЫХ ДАННЫХ ДЛЯ СИСТЕМЫ ТЕНДЕРОВ НА ОСНОВЕ ТЕХНОЛОГИИ БЛОКЧЕЙН[1]

Д. О. Кондырев

Институт математики им. С. Л. Соболева СО РАН, г. Новосибирск, Россия
Новосибирский государственный университет, г. Новосибирск, Россия
Лаборатория криптографии JetBrains Research, г. Новосибирск, Россия

На основе открытой блокчейн-платформы Ethereum разработана система тендеров, которая позволяет скрывать информацию о заявках на этапе запроса предложений. Создан новый метод, позволяющий решить проблему приватности информации в открытых блокчейн-системах с использованием криптографического протокола доказательства с нулевым разглашением zk-SNARK. Предложенный метод реализован в виде криптографической схемы на основе библиотеки libsnark. Для интеграции криптографической схемы в систему модифицирован Ethereum C++ клиент, куда добавлены новые функции и интерфейс для работы с ними в виде предкомпилированных контрактов.

Ключевые слова: *тендеры, распределенные системы, блокчейн, доказательство с нулевым разглашением, zk-SNARK, платформа Ethereum.*

DOI 10.17223/20710410/48/6

DEVELOPMENT OF A METHOD FOR HIDING PRIVATE DATA FOR A BLOCKCHAIN-BASED TENDER SYSTEM

D. O. Kondyrev

Sobolev Institute of Mathematics, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
Laboratory of Cryptography JetBrains Research, Novosibirsk, Russia

E-mail: dkondyrev@gmail.com

A tender system has been developed based on the Ethereum open blockchain platform that allows to hide the information about applications at the request for proposals stage. A new method has been created to solve the problem of information privacy in open blockchain systems using the zk-SNARK, cryptographic zero-knowledge proof protocol. The proposed method has been implemented as a cryptographic scheme based on the libsnark library. To integrate the cryptographic scheme into the system, the Ethereum C++ client has been modified — a new tenderzkp module has been

[1]Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации № 075-15-2019-1613, и лаборатории криптографии JetBrains Research.

added. It implements functions for creating and verifying zk-SNARK proofs. Interaction with the implemented cryptographic scheme from the smart contract codes is carried out through the new added precompiled contracts. A Solidity library has been created to work with these contracts. The JSON-RPC API of the Ethereum C++ client has been expanded to enable to call methods of the cryptographic scheme from third-party applications.

Keywords: *tenders, distributed systems, blockchain, zero-knowledge proof, zk-SNARK, Ethereum platform.*

Введение

На сегодняшний день большинство конкурсных закупок и электронных торгов проводится через специализированные информационные системы. Для таких систем критичным является вопрос доверия оператору торговой площадки. Участники должны быть уверены в том, что никто не имеет возможности нарушить правила проведения тендера или получить доступ к конфиденциальной информации. В рассмотренных системах вероятность нарушения этих правил не может быть полностью исключена.

Решить проблему доверия при проведении тендеров позволяет блокчейн-технология надёжного распределённого хранения записей о транзакциях. Преимущество этой технологии в том, что она позволяет взаимодействовать участникам напрямую без посредника — оператора площадки. При этом данные хранятся распределённо на узлах блокчейн-сети, история транзакций не может быть изменена или удалена [1 – 3].

Однако при использовании этой технологии данные сохраняются в открытом виде и доступны всем участникам, что не всегда приемлемо при создании промышленных программных систем. В случае с тендерами открытость информации нарушает тайну заявок, которая должна быть сохранена до окончания этапа запроса предложений. Это не позволяет проводить конкурсные закупки в существующих открытых блокчейн-системах.

Целью данной работы является разработка открытой блокчейн-системы для проведения тендеров, которая решила бы проблему приватности информации.

В работе проведён анализ предметной области, представлен краткий обзор технологий, рассмотрены существующие проблемы электронных торговых площадок и предложен новый метод сокрытия приватной информации в открытых блокчейн-системах для реализации конкурсных закупок. Разработанный метод основан на протоколе доказательства с нулевым разглашением zk-SNARK (zero-knowledge Succinct Non-Interactive Argument of Knowledge) и позволяет скрывать конфиденциальную информацию на этапе подачи заявок.

Для реализации предложенного метода модифицирован Ethereum C++ клиент, в который интегрирована разработанная криптографическая схема на основе библиотеки libsnark. Добавлены новые предкомпилированные контракты для работы с криптографической схемой и реализована Solidity-библиотека для работы с ними.

1. Основные требования к системе

Основными принципами процедуры проведения тендеров являются открытость, прозрачность, конкурентность, равенство участников и справедливость [4]. Исходя из этого, можно сформулировать требования, которым должна удовлетворять информационная система тендеров:

- T1 **Невозможность изменения информации.** Вся история транзакций в системе должна быть неизменяемой. Участники не должны иметь возможности исправлять данные зарегистрированных заявок, а организатор — изменять правила или результаты после окончания конкурса. Однако после того, как тендер опубликован, часто возникают уточнения или изменения. Все такие правки должны оформляться и регистрироваться в системе в виде отдельных документов.
- T2 **Невозможность раннего вскрытия заявок.** Ни у кого из пользователей системы (в том числе и у организатора тендера) не должно быть возможности просматривать данные предложений участников конкурса до завершения периода приема заявок.
- T3 **Анонимность заявок.** Участники не должны знать, кто подал заявки на тендер, до его завершения.
- T4 **Сокрытие факта подачи заявки.** Факт подачи конкретным пользователем заявки на какой-либо тендер должен быть скрыт от других пользователей системы, поскольку знание этого факта раскрывает информацию о деятельности пользователя и может нарушать принцип конкурентности.
- T5 **Запрет подмены пользователя.** Осуществление действий в системе (объявление тендера, подача заявки и др.) от имени другого пользователя должно быть запрещено.
- T6 **Открытость информации.** Вся информация должна быть в открытом доступе. Во-первых, это касается информации об объявленных тендерах, все потенциальные участники должны иметь к ней доступ, причём получать его одновременно, чтобы не нарушить принцип честности. Во-вторых, после завершения тендера сторонние наблюдатели должны иметь возможность проверить честность проведения конкурса, поэтому им необходим доступ к результатам, заявкам и всей истории операций.
- T7 **Невозможность нарушения сроков.** Заявки не могут быть поданы до начала процедуры подачи заявок и после её окончания.
- T8 **Гарантия выполнения правил тендера.** Все правила должны быть чётко зафиксированы и обязательны к исполнению всеми участниками процесса. Недопустимо изменение правил тендера после его объявления.
- T9 **Доказательство подачи заявки.** Участники конкурса должны иметь возможность доказать факт подачи своей заявки. При этом ни один пользователь системы не может подделать такое доказательство.

2. Обзор предлагаемого решения

Большинство существующих систем для проведения тендеров имеют общую схему функционирования. Какая-либо организация, выступающая в роли оператора, предоставляет площадку, которой могут пользоваться заинтересованные компании. В этом случае всё взаимодействие участников закупок с площадкой основано на доверии организации-оператору. Подобные платформы не удовлетворяют требованиям открытости и прозрачности и не всегда могут считаться надёжными системами.

В настоящее время ведутся исследования в области децентрализованных систем проведения тендеров. Технология блокчейн позволяет создать площадку, с помощью которой пользователи могут проводить тендеры и заключать договоры напрямую без участия посредника. Вся информация о тендерах хранится на всех узлах блокчейн-сети, что делает систему более отказоустойчивой. Кроме того, любая транзакция, ко-

торая записана в блокчейне, не может быть изменена или удалена. Корректность выполнения правил участниками контролируется смарт-контрактами, что позволяет избавиться от влияния человеческого фактора на результаты тендера (никто не может нарушить процедуру, поскольку все ограничения реализованы в виде программного кода, который не может быть изменён) [5].

Одним из достоинств технологии блокчейн является открытость информации. Любой пользователь всегда имеет возможность просмотреть информацию, хранящуюся в блокчейне, а также проследить всю историю транзакций. Но полная открытость всей информации нарушает требования к организации конкурсных закупок. Процедура проведения тендеров предполагает, что участники не имеют возможности ознакомиться с заявками других претендентов на стадии приёма заявок. Организатор тендера тоже должен получать доступ к заявкам только после того, как завершён их прием. Поэтому открытые блокчейн-системы для проведения тендеров не обеспечивают необходимый уровень приватности информации.

Решением проблемы приватности информации в блокчейн-системе может быть защита данных посредством шифрования. Согласно такому подходу, при подаче заявки участник генерирует симметричный ключ, шифрует информацию о заявке этим ключом и отправляет шифртекст в качестве своей заявки в блокчейн. После окончания срока приёма заявок все участники, отправившие заявки на тендер, должны отправить ключи, которыми эти заявки шифровались. Имея ключ и зашифрованную заявку, любой желающий может проверить корректность данных. Такая система предложена в [6]. Однако такой подход не позволяет проверить корректность зашифрованной заявки в момент её подачи. Ещё одним недостатком является то, что все участники могут наблюдать факт подачи заявки пользователем.

Систему тендеров предлагается реализовать на основе технологии блокчейн, потому что она позволяет обеспечить прозрачность и открытость процедуры проведения конкурса; проверка действий участников может быть реализована в виде смарт-контрактов, которые выступают гарантом выполнения правил.

Необходимо реализовать возможность анонимной подачи заявок на тендеры и добавить сокрытие информации о заявках, при этом обеспечив проверку её корректности. По истечении сроков подачи заявок информация должна раскрываться и сохраняться в открытом виде. Таким образом, вся история будет открытой, что позволит обеспечить прозрачность процедуры тендеров, при этом не будут нарушаться правила конкурсных закупок.

3. Проблема сокрытия информации

На сегодняшний день можно выделить два основных подхода к сокрытию информации о транзакциях в блокчейн-сети [7]:

— механизм смешивания (mixing);
— подход на основе доказательства с нулевым разглашением.

3.1. Механизм смешивания

Протоколы, основанные на данном подходе, принимают разные формы, но все реализуют одну идею. Базовая перемешивающая сеть, также известная как mixnet, является протоколом маршрутизации, в котором сервер принимает в качестве входных сообщения от нескольких отправителей, перемешивает их и отправляет в случайном порядке получателям. Цель такой сети — исключить возможность отследить соответствия между отправителями и получателями транзакций [7].

Такой подход реализуется в CoinShuffle [8], XIM [9], Mixcoin [10] и многих других системах и протоколах. Данный подход обладает рядом недостатков, которые не позволяют применять его в тендерных системах:

1) лишь частично решается проблема анонимности, поскольку не полностью скрывается информация о транзакциях;
2) смешивание применимо только для задачи анонимизации транзакций (позволяет скрыть отправителя); нет возможности расширить алгоритм для более общего случая сокрытия произвольных данных в блокчейн-транзакциях.

3.2. Доказательство с нулевым разглашением

Доказательство с нулевым разглашением — криптографический протокол, в котором принимают участие две стороны — доказывающая и проверяющая (верификатор). Цель протокола заключается в том, чтобы верификатор мог убедиться, что доказывающая сторона обладает знанием секретного параметра. При этом сам секретный параметр не должен раскрываться верификатору или кому-либо ещё [11].

Это может быть представлено в виде программы с двумя входами $C(x, a)$. Вход x является открытым, a — секретный параметр (witness). Выход программы бинарный (`true` либо `false`). Задаётся конкретный общедоступный x. Задача состоит в том, чтобы доказать, что доказывающая сторона знает секретный параметр a, такой, что $C(x, a) = \texttt{true}$.

Доказательство с нулевым разглашением по определению должно удовлетворять следующим трём свойствам:

1) Полнота: если утверждение верно и обе стороны следуют одному и тому же протоколу, то верификатор может убедиться в истинности утверждения.
2) Устойчивость: если утверждение ложно, верификатор с большой вероятностью не будет убеждён в его истинности.
3) Нулевое разглашение: верификатор не получает дополнительной информации.

Концепция интерактивных систем доказательства с нулевым разглашением впервые введена в работе [12]. За годы исследований в области доказательства с нулевым разглашением системы, основанные на этом методе, постепенно улучшались с упором на оптимизацию их эффективности для конкретных приложений. Это привело к появлению алгоритмов, которые существенно сократили количество раундов взаимодействия участников протокола.

Особенности технологии блокчейн, которая взята за основу построения системы тендеров, накладывают ряд ограничений на используемые криптографические протоколы, в частности на доказательство с нулевым разглашением. Поскольку блокчейн является распределённой системой, пользователи могут не быть в сети одновременно. При этом доказательство должно быть доступно всем участникам. После того как доказательство предоставлено, любой пользователь должен иметь возможность проверить его корректность в любой момент времени. Это делает применение интерактивных протоколов доказательства с нулевым разглашением в блокчейн-системах труднореализуемым.

В работе [13] впервые предложен неинтерактивный протокол доказательства с нулевым разглашением. Неинтерактивная система содержит только одно сообщение (доказательство), которое доказывающая сторона отправляет верификатору, т. е. взаимодействие между участниками протокола сводится к одному раунду. Дальнейшие исследования в области неинтерактивных протоколов были направлены на оптимизацию вычислительной эффективности и сокращение размера доказательства.

Существенным прорывом в этом направлении можно считать появление zk-SNARK [14], который сделал возможным эффективное использование неинтерактивных протоколов доказательства с нулевым разглашением в блокчейн-системах.

3.3. Криптографический протокол zk-SNARK

zk-SNARK — это криптографический протокол неинтерактивного доказательства знания с нулевым разглашением [15]. Он позволяет доказывать, что некоторые приватные данные удовлетворяют системе ограничений, выраженной в виде арифметической схемы C, не раскрывая эти данные.

zk-SNARK представляет собой тройку алгоритмов полиномиального времени выполнения (Gen, P, V):

- $Gen(\lambda, C) \to (pk, vk)$. Этот алгоритм принимает в качестве входных данных параметр безопасности λ и арифметическую схему C. На их основе генератор Gen создаёт пару ключей — ключ доказательства (pk, proving key) и ключ верификации (vk, verification key). Оба ключа публикуются как открытые параметры и могут использоваться любое количество раз для создания доказательства и проверки его корректности.
- $P(pk, x, a) \to \pi$. Принимая на вход ключ доказательства pk и любые (x, a), где x — публичные данные, a — секретный параметр, алгоритм P выводит неинтерактивное доказательство π.
- $V(vk, x, \pi) \to b$. Принимая на вход ключ верификации vk, публичные данные x и доказательство π, верификатор V выдаёт $b = 1$, если доказательство является корректным, и 0 иначе.

Данная конструкция удовлетворяет всем требованиям, предъявляемым к алгоритмам доказательства с нулевым разглашением [15].

Преимущество zk-SNARK над другими протоколами доказательства с нулевым разглашением заключается в гарантиях эффективности: длина доказательства зависит только от параметра безопасности, а время проверки не зависит от размера схемы и секретного параметра. Таким образом, zk-SNARK можно рассматривать как неинтерактивный протокол с коротким доказательством и быстрым временем верификации, что делает его наиболее подходящим для использования в блокчейн-системах [16].

3.4. Сокрытие информации в платформе Ethereum

Вся информация в Ethereum-блокчейне хранится в открытом виде, а транзакции не скрывают своих значений. Каждая транзакция содержит адреса аккаунтов отправителя и получателя и передаваемые данные [17]. При этом нет возможности средствами Ethereum скрыть часть полей транзакции (например, нельзя скрыть адрес аккаунта отправителя транзакции).

В zk-SNARK процедура проверки доказательства состоит из операций на эллиптических кривых. В частности, верификатор требует скалярного умножения и сложения на группе точек эллиптических кривых, а также вычислительно более сложной операции — билинейного спаривания.

Ethereum предоставляет реализацию этих операций в виде предварительно скомпилированных контрактов. С их помощью есть возможность реализовать схемы на основе доказательства с нулевым разглашением в коде смарт-контрактов [18, 19].

Сами алгоритмы генерации и верификации доказательства zk-SNARK не реализованы в платформе. В связи с этим возникает ряд проблем при использовании схем на основе zk-SNARK в Ethereum:

1) Нет возможности создавать сложные схемы. Все алгоритмы должны быть реализованы в смарт-контрактах, на размер кода которых накладываются жёсткие ограничения, а криптографические схемы, как правило, требуют большого количества операций.
2) Все алгоритмы приходится реализовывать вручную.
3) Для каждого нового контракта необходимо генерировать отдельные параметры.

3.5. Библиотека libsnark

Libsnark — криптографическая библиотека с открытым исходным кодом, написанная на языке C++, которая обеспечивает эффективные реализации конструкций zk-SNARK [20]. Библиотека является самым быстрым и полным набором доказательств с нулевым разглашением, доступных на данный момент [16].

Для создания криптографических схем zk-SNARK библиотека представляет набор высокоуровневых интерфейсов (gadgetlib1, gadgetlib2 и др.). Эти интерфейсы осуществляют преобразования высокоуровневых спецификаций в арифметические схемы, реализованные в ядре библиотеки. С их помощью можно строить новые криптографические схемы на основе реализованных низкоуровневых примитивов.

4. Архитектура системы проведения тендеров

В данной работе создана система проведения тендеров на основе платформы Ethereum. Архитектура системы состоит из следующих модулей, которые изображены на рис. 1:

— смарт-контракты в Ethereum, обеспечивающие работу с блокчейном;
— модифицированный Ethereum-клиент;
— Java–приложение, предоставляющее высокоуровневый интерфейс для работы с системой.

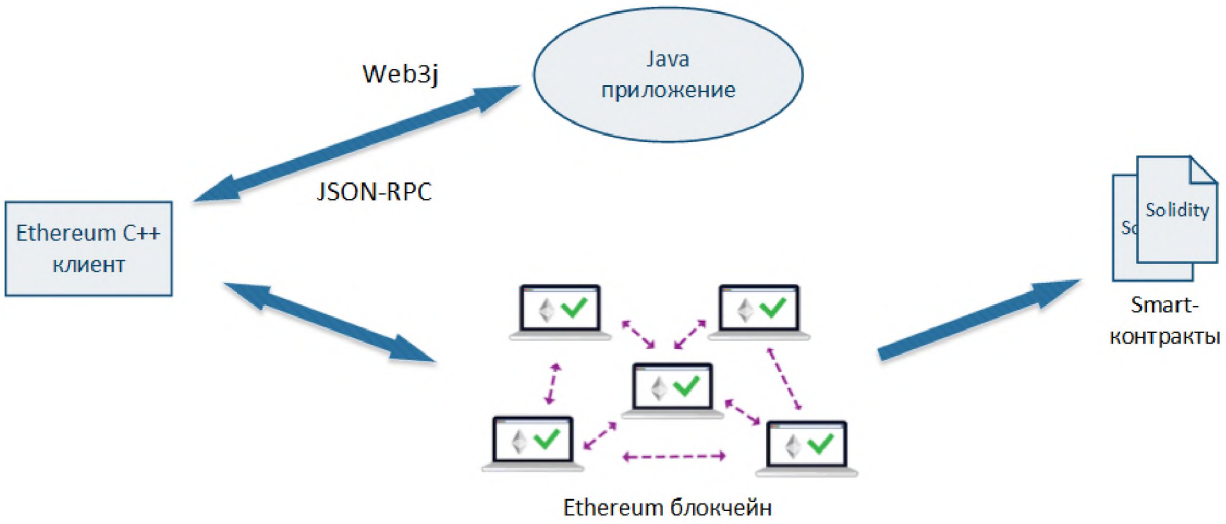


Рис. 1. Схема архитектуры системы

Ethereum-клиент представляет собой реализацию протокола Ethereum. Это программа, которая поддерживает состояние цепочки блоков транзакций и предоставляет API для проведения транзакций и запроса информации о текущем состоянии цепочки блоков. Через него происходит всё взаимодействие с блокчейном, в том числе со смарт-контрактами. Клиент включает в себя реализацию виртуальной машины

Ethereum, которая запускается при выполнении транзакций, взаимодействующих со смарт-контрактами [21]. Ещё одной его задачей является осуществление сетевого взаимодействия с другими клиентами, вместе они образуют единую блокчейн-сеть.

Существует несколько реализаций Ethereum-клиента на различных языках программирования. В разработанной системе используется C++ Ethereum-клиент (aleth), поскольку интегрировать библиотеку libsnark в код этого клиента оказалось наиболее просто и эффективно. При этом система не полагается на особенности реализации клиента aleth. Все необходимые модификации можно внести и в другие Ethereum-клиенты, для чего не потребуется менять архитектуру системы.

Всё взаимодействие между Java-приложением и Ethereum-клиентом (вызов методов смарт-контрактов, прослушивание событий, запрос информации) осуществляется с помощью JSON RPC. JSON-RPC — это легковесный протокол удалённого вызова процедур (RPC) без сохранения состояния [22]. Он использует JSON в качестве формата данных, а в качестве протокола передачи сообщений в системе используется HTTP.

Протокол JSON RPC, реализуемый клиентом Ethereum, является довольно низкоуровневым, и работать с ним напрямую неэффективно. Поэтому для взаимодействия со смарт-контрактами используется Java-библиотека web3j, которая работает поверх JSON-RPC API клиента Ethereum. Web3j позволяет работать с блокчейном без дополнительных накладных расходов на написание собственного интеграционного кода. Библиотека поддерживает все методы JSON-RPC API и может работать с любым клиентом Ethereum, который его реализует [23].

Для более удобного взаимодействия со смарт-контрактами web3j позволяет создать Java-оболочки. На основе кода смарт-контрактов порождаются классы, которые предоставляют функции создания и развертывания смарт-контракта, вызова его функций и выполнения транзакций из Java-кода [24].

Рассмотрим подробно, как устроены отдельные модули.

4.1. Модифицированный Ethereum-клиент

Реализация криптографических протоколов на основе zk-SNARK предполагает генерацию и проверку ограничений, а также выполнение операций над эллиптическими кривыми.

Изначально реализация этих операций была выполнена в виде смарт-контрактов, однако оказалась неэффективной. Код смарт-контрактов хранится в блокчейне, а его выполнение происходит при процессе проверки корректности транзакций в виртуальной машине Ethereum на каждом узле сети, поддерживающем цепочку блоков транзакций. Поэтому к коду смарт-контрактов предъявляются жёсткие требования по эффективности вычисления и размеру. А криптографические операции, необходимые для работы системы, являются вычислительно затратными и при реализации непосредственно в смарт-контрактах сильно увеличивают размер кода. Поэтому было решено реализовать этот протокол на стороне Ethereum-клиента.

4.2. Криптографическая схема

Для реализации алгоритма сокрытия информации о заявках на основе доказательства с нулевым разглашением в Ethereum C++ клиент добавлен отдельный модуль *tenderzkp*. Он построен на базе протокола zk-SNARK с предобработкой (preprocessing zk-SNARK) для NP-полного языка системы ограничений ранга 1 (R1CS — rank-1 constraint systems). Протокол использует эллиптическую кривую Баррето — Наерига. Реализация криптографической схемы предоставлена библиотекой libsnark [20].

Основной интерфейс этого модуля составляют две функции:

— $generate\_proof(proving\_key,\ public\_input,\ private\_input) \to proof$;
— $verify\_proof(verification\_key,\ public\_input,\ proof) \to \{true, false\}$.

Функция генерации доказательства $generate\_proof$ принимает на вход открытые ($public\_input$) и приватные ($private\_input$) данные, а также ключ доказательства ($proving\_key$). Приватными данными для заявки на тендер являются:

— ID участника, подающего заявку;
— ID тендера, на который подается заявка;
— время подачи заявки;
— сумма предложения.

К открытым данным относится информация о тендерах (ID тендеров, время окончания приёма заявок, максимально допустимые суммы предложений), которые проводятся на данный момент, и ID пользователей системы.

Для приватных данных заявки необходимо проверить, что выполнены следующие условия:

— сумма предложения не превышает максимально допустимой для тендера, на который подаётся заявка;
— время подачи заявки не превышает времени окончания приёма заявок на данный тендер;
— участник с таким ID действительно зарегистрирован в системе.

Чтобы создать доказательство с нулевым разглашением, необходимо выразить эти условия в виде ограничений на приватные и открытые входные данные. Для этого создан класс $TenderGadget$, в котором реализована криптографическая схема.

$TenderGadget$ выражает условия корректности заявки с помощью базовых схем библиотеки gadgetlib1. Для этого используются функции $comparison\_gadget$, $conjunction\_gadget$ и $disjunction\_gadget$, реализующие сравнение целочисленных значений (в данном случае это ID тендеров, ID пользователей, суммы предложений и время) и логические операции конъюнкции и дизъюнкции соответственно.

Построенная арифметическая схема преобразуется в более низкоуровневый вид — систему ограничений ранга 1. Полученное R1CS-представление используется в дальнейшем алгоритмами генерации и верификации доказательства libsnark.

На основе R1CS-представления генерируется доказательство утверждения, что входные данные удовлетворяют системе ограничений. Это доказательство является возвращаемым значением функции $generate\_proof$.

Функция проверки доказательства $verify\_proof$ принимает открытые данные ($public\_input$), доказательство, сгенерированное функцией $generate\_proof$ ($proof$), и ключ верификации ($verification\_key$). Она возвращает `true`, если доказательство корректно, и `false` иначе.

Пара ключей (доказательства и верификации), необходимая для работы алгоритмов zk-SNARK, является общей для всей схемы, т. е. для всех контрактов тендеров используются одни и те же ключи. Благодаря этому, есть возможность не хранить пару ключей в смарт-контрактах, а переложить функцию управления ими на Ethereum-клиент, что более эффективно как с точки зрения используемой памяти, так и с точки зрения скорости загрузки. В разработанной системе ключи передаются в Ethereum-клиент в виде конфигурационных файлов и загружаются модулем $tenderzkp$ при вызовах функций $generate\_proof$ и $verify\_proof$.

В разработанной схеме на тестовых данных параметры имеют следующие размеры:

— параметр безопасности λ — 192 байта;
— ключ доказательства pk — 231 кбайт;
— ключ верификации vk — 1 884 байта;
— публичные данные x — 480 байт;
— секретный параметр a — 192 байта;
— доказательство π — 576 байт;
— арифметическая схема C задаётся в коде (класс *TenderGadget*). Её размер в преобразованном для алгоритмов доказательства и верификации виде составляет около 77 кбайт.

4.3. Взаимодействие с криптографической схемой

У смарт-контрактов должна быть возможность взаимодействовать с реализованной криптографической схемой — вызывать функции генерации и проверки доказательства и получать возвращаемые значения. Выполнение кода смарт-контрактов происходит в виртуальной машине Ethereum, поэтому одним из возможных вариантов реализации взаимодействия было бы добавление новых операций в EVM. Но при таком решении необходимо вносить большое количество изменений в платформу Ethereum — не только дополнить набор команд EVM, но и внести соответствующие доработки в компиляторы высокоуровневых языков написания смарт-контрактов (таких как Solidity).

Альтернативным подходом является создание предкомпилированных контрактов. Предкомпилированный контракт — это смарт-контракт, который имеет фиксированный адрес и код которого реализован непосредственно в Ethereum-клиентах. Большая часть криптографических операций в Ethereum (восстановление адреса аккаунта из ECDSA подписи, хеш-функции SHA-256 и RIPEMD-160 и др.) реализована именно в виде предкомпилированных контрактов [25]. Такие контракты являются тестовыми изменениями архитектуры, которые впоследствии могут стать частью протокола Ethereum [26].

В разработанной системе решено использовать второй подход, чтобы минимизировать количество изменений относительно существующих реализаций протокола Ethereum. В Ethereum C++ клиент добавлены новые предкомпилированные контракты с адресами `0x00...09` и `0x00...0a`. При обращении к ним из кода смарт-контрактов вызываются функции *generate_proof* и *verify_proof* добавленного модуля *tenderzkp*.

Генерация доказательства должна происходить вне блокчейна, так как приватная информация заявки не должна попасть в открытый доступ на данном этапе. Всё взаимодействие с клиентом происходит через JSON-RPC API, поэтому чтобы добавить возможность вызывать методы криптографической схемы из сторонних приложений, добавлены соответствующие интерфейсы в модуль *web3jsonrpc* Ethereum C++ клиента. На рис. 2 представлена схема всех модификаций, внесённых в Ethereum C++ клиент.

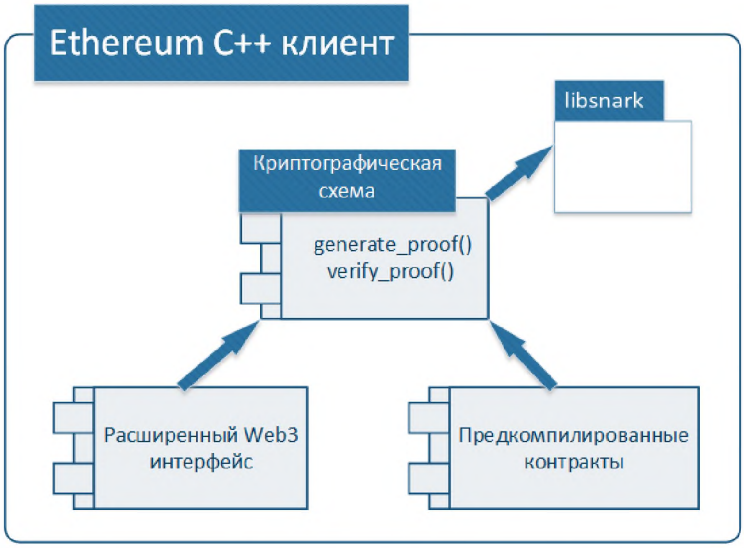


Рис. 2. Модификации Ethereum-клиента, реализованные в системе

4.4. Смарт-контракты

Смарт-контракты — это объекты в блокчейне, которые содержат своё состояние и код функций. Они написаны на статически типизированном высокоуровневом языке программирования Solidity, предоставляемом платформой Ethereum. Общая схема модуля смарт-контрактов изображена на рис. 3.

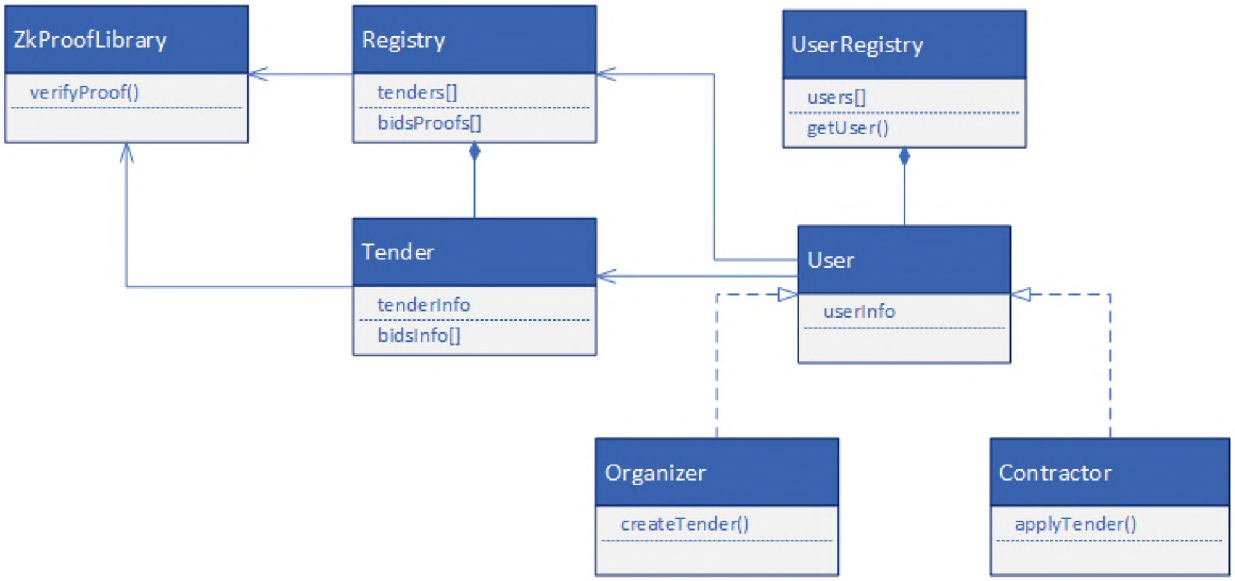


Рис. 3. Схема модуля смарт-контрактов

Основным смарт-контрактом является реестр тендеров (Registry). В нём хранится информация о зарегистрированных тендерах и ссылки на их контракты, а также все закрытые заявки на тендеры (доказательства, полученные алгоритмом криптографической схемы). Через этот контракт проходит регистрация всех тендеров в системе, а также подача закрытых заявок. Закрытая заявка может быть зарегистрирована в реестре только в том случае, если проходит проверка доказательства, реализованная в предкомпилированном контракте.

Для каждого тендера создается отдельный контракт (Tender), в котором сохраняется информация о нём, текущая стадия конкурса, поданные открытые заявки и информация о результатах. Подача открытых заявок происходит через контракт тендера. При этом для каждой открытой заявки происходит проверка корректности — вызывается предкомпилированный контракт, который генерирует доказательство на основе приватных данных, и полученное доказательство сравнивается с ранее зарегистрированным в реестре.

Пользователи работают с контрактами реестра и тендеров не напрямую, вызывая их методы, а через специальные смарт-контракты пользователей (User, Organizer, Contractor). Это позволяет реализовать разграничение прав пользователей (заказчик может объявлять новые тендеры, а участник конкурса может только подавать заявки на существующие). Контракт хранит информацию о пользователе и адрес Ethereum-аккаунта, к которому он привязан. Только транзакции, отправленные от имени этого аккаунта, считаются корректными, поэтому никто, кроме владельца аккаунта, не может совершать действия от имени этого контракта. Кроме того, для каждого заказчика в контракте сохраняется список объявленных им тендеров, а для каждого участника конкурса — все поданные заявки.

Информация обо всех пользователях хранится в реестре пользователей (UserRegistry). Этот контракт регулирует добавление новых пользователей в систему, а также смену Ethereum-аккаунтов, от имени которых работают пользователи.

Стоит отметить, что реализация всех действий пользователей в системе через специальные контракты User вместе с возможностью смены аккаунта позволяют решить одну из фундаментальных проблем блокчейн-систем — потерю приватного ключа аккаунта. Она заключается в том, что отсутствует возможность восстановления приватного ключа аккаунта, и при его потере пользователь не сможет использовать этот аккаунт для дальнейшей работы в блокчейн-системе. В разработанной системе в случае потери ключа есть возможность сменить аккаунт, не изменяя контракт пользователя. При этом информация о пользователе и история его действий в системе (создании тендеров и подаче заявок) сохраняются.

Использование проверок в коде функций смарт-контрактов исключает возможность нарушения участниками правил проведения конкурсных закупок, таких, как:

— объявление тендеров от имени другого пользователя;
— объявление победителем участника, заявка которого не была зарегистрирована, и т.д. [27].

Все проверки, которые осуществляются в смарт-контрактах, гарантированно выполняются, так как блокчейн-транзакции, которые не удовлетворяют условиям в коде проверок, откатываются. Для реализации проверок используется стандартная функция языка Solidity *require*().

Функция *require*() компилируется в набор инструкций, которые осуществляют проверку условия, и инструкцию REVERT (`0xfd`) виртуальной машины Ethereum, к которой переходит управление в случае, когда условие не выполняется. Если во время выполнения кода смарт-контракта в виртуальной машине встречается инструкция REVERT, выполнение кода останавливается, а все изменения, произведенные транзакцией, отменяются. Такой откат всех изменений позволяет сохранить атомарность транзакции. При этом сама транзакция сохраняется в блокчейне.

Контракты являются частью протокола и утверждаются участниками до старта работы системы.

Во время работы системы могут создаваться новые объекты контрактов двух типов — User и Tender. Однако их создание производится не напрямую пользователями, а через контракты UserRegistry и Registry соответственно, т. е. сам код контрактов User и Tender предварительно скомпилирован и интегрирован в код контрактов UserRegistry и Registry. У пользователей отсутствует возможность добавлять свои собственные реализации каких-либо контрактов.

Если в процессе работы обнаруживаются ошибки в смарт-контрактах, они могут быть исправлены в новых версиях этих же контрактов. Переход на новую версию осуществляется только в случае одобрения изменений участниками системы.

Взаимодействие с предкомпилированными смарт-контрактами не может быть реализовано средствами языка Solidity, для вызова кода функций таких контрактов используются ассемблерные вставки. Для удобства работы создана Solidity-библиотека. Она инкапсулирует низкоуровневое взаимодействие с предкомпилированными контрактами и предоставляет интерфейс для работы с ними в виде Solidity-функций.

5. Алгоритм работы системы

Рассмотрим более подробно алгоритм работы системы. Можно выделить следующие этапы проведения тендера:

1) Заказчик создаёт контракт Tender, в котором размещает всю необходимую информацию о проводимом конкурсе.
2) Пользователи подают скрытые заявки в общий реестр тендеров.
3) После окончания срока приёма заявок пользователи вскрывают заявки (отправляют открытую информацию в контракт Tender со ссылкой на скрытую заявку). Все заявки, которые не были вскрыты, аннулируются.
4) После окончания срока предоставления открытой информации заказчик оценивает заявки и определяет победителя.

Отдельные крупные этапы данного алгоритма — подача скрытой заявки и раскрытие информации.

5.1. Процесс подачи скрытой заявки

Процесс подачи скрытой заявки, схематично представленный на рис. 4, проходит следующим образом:

1) Пользователь создаёт новый анонимный Ethereum-аккаунт.
2) После этого пользователь формирует заявку на выбранный тендер, которая содержит необходимые приватные данные (ID участника, ID тендера, текущее время, сумму предложения).
3) На основе приватных данных заявки создаётся публичное доказательство. Для этого через JSON-RPC API модифицированного Ethereum-клиента вызывается функция генерации доказательства криптографической схемы доказательства с нулевым разглашением.
4) Далее от имени анонимного аккаунта пользователь отправляет доказательство в контракт Registry.
5) Смарт-контракт Registry осуществляет проверку корректности доказательства, вызывая код предкомпилированного контракта верификации. Если проверка пройдена успешно, заявка записывается в хранилище контракта Registry. В противном случае заявка считается некорректной и отклоняется.

Благодаря тому, что для каждой подачи заявки генерируется новый аккаунт, нельзя отследить, кто именно записывает публичное доказательство. Это обеспечивает со-

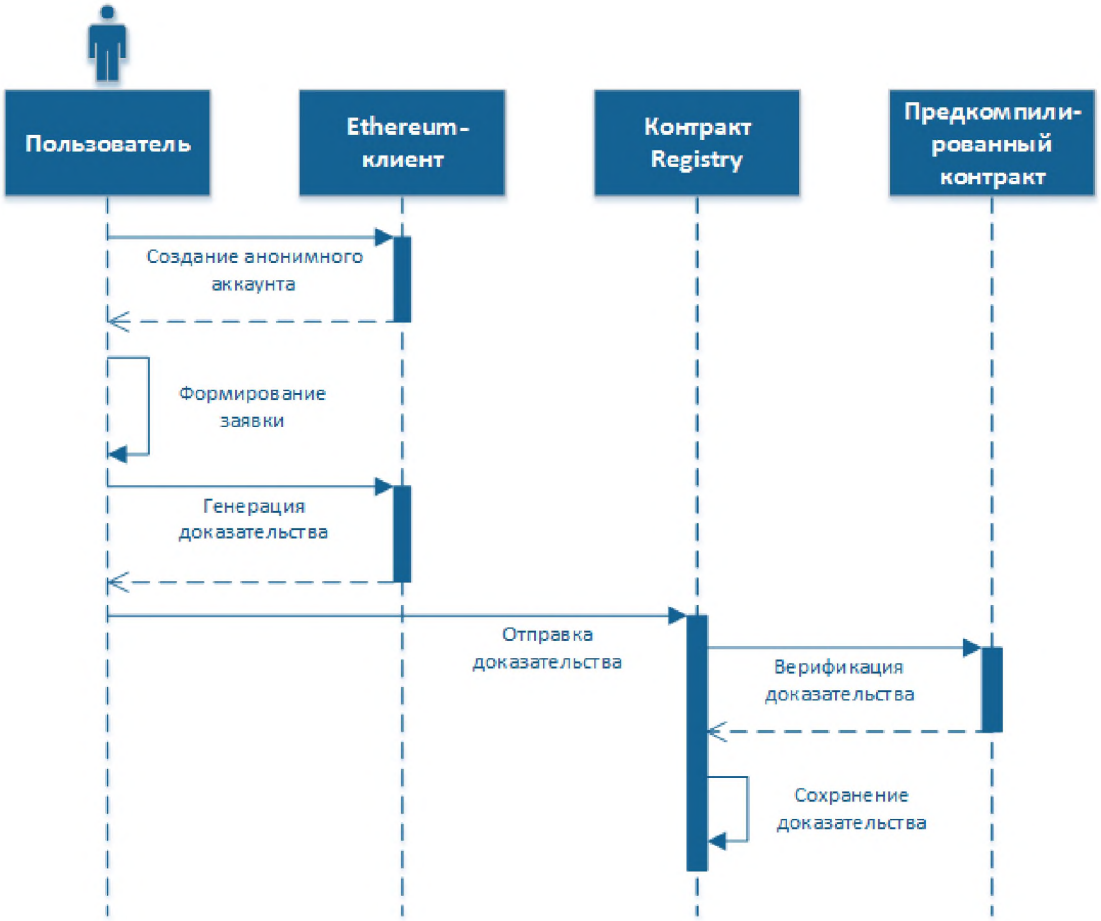


Рис. 4. UML-диаграмма последовательности процесса подачи скрытой заявки

крытие не только приватных данных заявки, но и самого факта подачи пользователем заявки на конкретный тендер.

Стоит отметить, что первые три шага описанного алгоритма выполняются вне блокчейна, поэтому секретная информация не видна никому, кроме самого пользователя.

5.2. Процесс раскрытия заявки

После окончания срока приёма заявок начинается этап раскрытия информации (рис. 5). На этом этапе:

1) Пользователь отправляет раскрытые данные в контракт Tender, указывая ссылку на закрытую заявку.
2) Контракт тендера запрашивает из контракта Registry зарегистрированное доказательство.
3) Далее вызывается предкомпилированный контракт, который на основе раскрытых приватных данных генерирует доказательство.
4) После этого контракт Tender производит ряд проверок:
   — Предоставленная информация соответствует ранее зарегистрированной в реестре закрытой заявке. Для этого производится сравнение сгенерированного доказательство с сохранённым в реестре. Их совпадение свидетельствует о том, что доказательства сгенерированы на основе одних и тех же приватных данных и заявка является подлинной. Если они не сов-

падают, это означает, что доказательство, соответствующее подаваемой открытой заявке, не было зарегистрировано на этапе запроса предложений, и такая заявка, согласно правилам проведения тендеров, не может принимать участие в конкурсе.
   - ID пользователя в заявке совпадает с ID пользователя, который отправляет данные.
   - ID тендера в заявке совпадает с ID тендера контракта.

5) При успешном прохождении проверок предоставленная информация сохраняется в контракте.

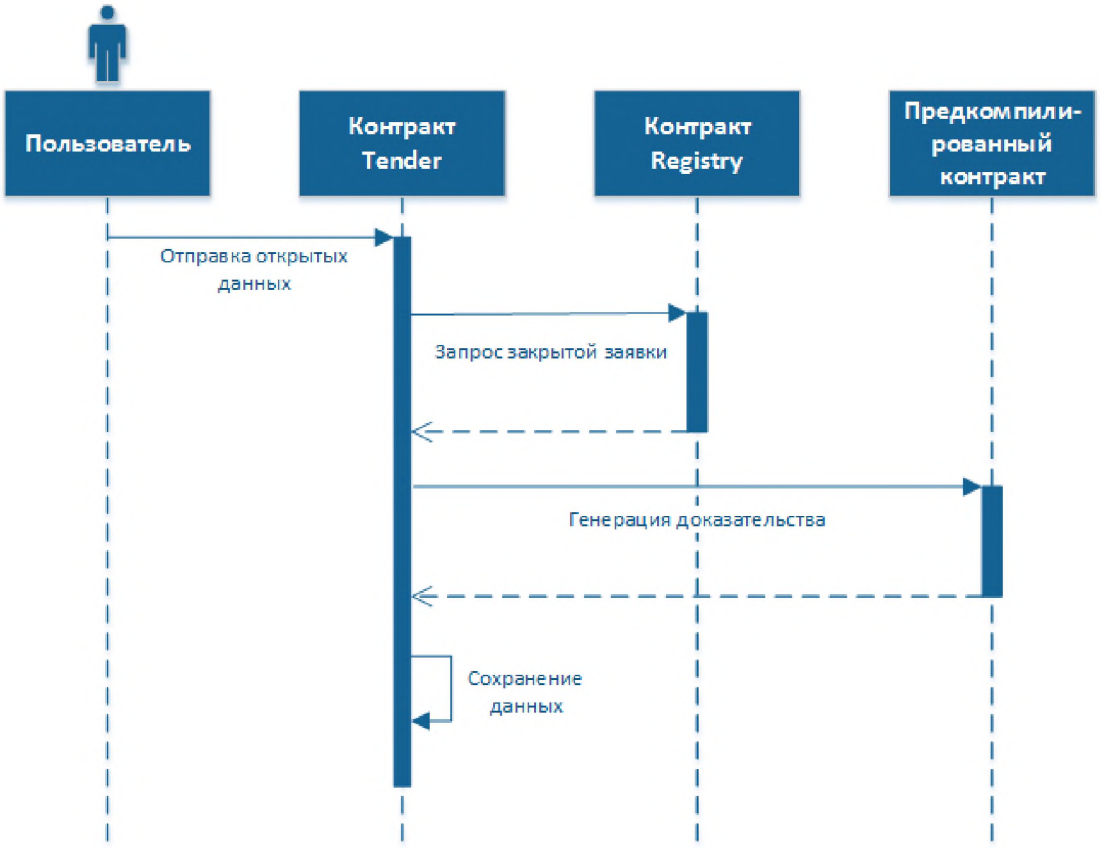


Рис. 5. UML-диаграмма последовательности процесса раскрытия заявки

Стоит отметить, что при раскрытии заявок пользователь производит действия в блокчейн-сети от имени своего аккаунта. Это позволяет удостовериться в личности пользователя и исключить возможность подачи заявки от лица другого участника.

Алгоритм работы системы, в основе которого лежит протокол доказательства с нулевым разглашением, кроме возможности проверять корректность информации при подаче заявки, даёт ещё несколько преимуществ.

Во-первых, уже на этапе подачи заявок можно собирать статистику, которая может быть использована для различных целей.

Во-вторых, при раскрытии заявок можно предоставлять в открытом виде только некоторые данные, гарантируя корректность всей остальной информации.

В-третьих, есть возможность реализовать алгоритм предоставления информации только отдельным участникам на этапе вскрытия заявок. Такой алгоритм может быть применён в системах закрытых тендеров, где информация о раскрытых заявках долж-

на быть доступна только организатору закупки. При этом за счёт использования алгоритмов доказательства с нулевым разглашением все остальные участники смогут проверить корректность поданных заявок и удостовериться, что победитель тендера выбран согласно правилам, не получая доступ к самой информации заявок.

Таким образом, реализованный алгоритм даёт более богатые возможности для расширения функциональности системы.

6. Развёртывание системы

Развёртывание любой программной системы, созданной на базе платформы Ethereum, может быть осуществлено двумя основными способами:

— в основной блокчейн-сети Ethereum (Ethereum Mainnet);
— в собственной блокчейн-сети.

Разработанный метод сокрытия приватной информации полагается на модификации, внесенные в Ethereum-клиент, а внедрение модифицированных Ethereum-клиентов в основную сеть невозможно из-за несовместимости протокола с обычными узлами. Вызов предкомпилированных контрактов будет невозможен на обычных узлах, а механизм консенсуса полагается на то, что все операции должны одинаково выполняться на всех узлах блокчейн-сети. Поэтому для системы проведения тендеров необходимо развернуть собственную блокчейн-сеть, состоящую из узлов, поддерживающих внесённые модификации. Это могут быть модифицированные Ethereum C++ клиенты, которые описаны ранее, либо любые другие клиенты, в которых добавлены предкомпилированные контракты с теми же адресами и реализующие описанную криптографическую схему доказательства с нулевым разглашением.

Поскольку предполагается использовать отдельную блокчейн-сеть, механизм оплаты действий в системе может регулироваться сообществом участников путём определения своих правил поверх существующей стандартной схемы оплаты транзакций в Ethereum. Например, взимание комиссии можно полностью отменить. Кроме того, в отдельной блокчейн-сети количество транзакций должно быть значительно меньше, чем в основной сети Ethereum, что увеличивает пропускную способность системы.

Для работы криптографической схемы необходима пара ключей (доказательства и верификации). Их генерация должна выполняться доверенной стороной. Получившиеся в результате открытые параметры публикуются и становятся доступными для всех сторон. В разработанной системе ключи передаются в качестве конфигурационных файлов при старте Ethereum-клиента. Процесс генерации ключей выполняется только один раз, после этого доверенная сторона не требуется.

Эта фаза является критической с точки зрения безопасности системы. Любой, кто обладает параметром безопасности, на основе которого сгенерированы ключи, получит возможность генерировать ложные доказательства, которые будут приняты алгоритмом верификации как корректные. Поэтому при внедрении системы процедуре генерации стоит уделить особое внимание. Как правило, используются многосторонние протоколы для безопасной генерации параметров, которые дают возможность не полагаться на честность единственного участника. В процессе инициализации параметров участвует ряд сторон, использующих многосторонний протокол для генерации ключей доказательства и верификации. Для обеспечения надёжности созданной криптографической схемы достаточно, чтобы хотя бы одна из сторон была честна. Распределённый протокол генерации параметров для zk-SNARK приведён в [16].

Заключение

Предложена и реализована система тендеров, которая удовлетворяет критериям безопасности, открытости и конфиденциальности. Вопрос доверия решён с помощью технологии блокчейн, а сокрытие приватной информации — с помощью алгоритмов доказательства с нулевым разглашением.

Разработан принципиально новый метод, позволяющий решить проблему приватности информации в блокчейн-системах с использованием алгоритмов доказательства с нулевым разглашением. Метод позволяет участникам зафиксировать факт подачи заявки на тендер, не раскрывая её содержания.

Предложенный и реализованный метод может быть использован не только для тендеров, но и в других системах, где есть необходимость скрывать часть информации в открытой блокчейн-сети. Он расширяет область применения технологии блокчейн в промышленных программных комплексах.

ЛИТЕРАТУРА

1. *Wattenhofer R.* The Science of the Blockchain. 1st ed. Inverted Forest Publishing, 2016. 115 p.
2. *Nakamoto S.* Bitcoin: A Peer-to-Peer Electronic Cash System. `https://bitcoin.org/bitcoin.pdf`
3. *Равал С.* Децентрализованные приложения. Технология Blockchain в действии. СПб.: Питер, 2017. 192 с.
4. *Кузнецов К. В.* Конкурентные закупки: торги, тендеры, конкурсы. СПб.: Питер, 2005. 368 с.
5. *Кондырев Д. О., Бобров В. С., Ефремов И. Е., Власов В. Н.* Система проведения тендеров на основе платформы Ethereum // Вестник НГУ. Сер. Информационные технологии. 2017. Т. 15. № 3. С. 31–39.
6. *Hardwick F. S., Akram R. N., and Markantonakis K.* Fair and transparent blockchain based tendering framework — A step towards open governance // IEEE Intern. Conf. TrustCom/BigDataSE, New York, USA, 2018. P. 1342–1347.
7. *Heilman E., Baldimtsi F., and Goldberg S.* Blindly signed contracts: Anonymous on-blockchain and off-blockchain bitcoin transactions // Intern. Conf. Financial Cryptography and Data Security. Springer, 2016. P. 43–60.
8. *Ruffing T., Moreno-Sanchez P., and Kate A.* Coinshuffle: practical decentralized coin mixing for bitcoin // ESORICS 2014. LNCS. 2014. V. 8713. P. 345–364.
9. *Bissias G., Ozisik A. P., Levine B. N., and Liberatore M.* Sybil-resistant mixingfor bitcoin // Proc. WPES'14. Scottsdale, Arizona, USA, November 2014. P. 149–158.
10. *Bonneau J., Narayanan A., Miller A., et al.* Mixcoin: anonymity for bitcoin with accountable mixes // Intern. Conf. Financial Cryptography and Data Security. Springer, 2014. P. 486–504.
11. *Шнайер Б.* Прикладная криптография: протоколы, алгоритмы и исходные коды на языке С. 2-е изд. СПб.: ООО «Альфа-книга», 2018. 1040 с.
12. *Goldwasser S., Micali S., and Rackof C.* The knowledge complexity of interactive proof systems // STOC'85. Proc. 17th Ann. ACM Symp. Theory of Computing. Providence, Rhode Island, USA, 1985. P. 291–304.
13. *Blum M., Feldman P., and Micali S.* Non-interactive zero-knowledge proof systems and applications // STOC'88. Proc. 20th Ann. ACM Symp. Theory of Computing. Chicago, USA, 1988. P. 103–112.
14. *Ben-Sasson E., Chiesa A., Genkin D., et al.* SNARKs for C: Verifying program executions succinctly and in zero knowledge // CRYPTO'2013. LNCS. 2013. V. 8043. P. 90–108.

15. *Ben-Sasson E., Chiesa A., Garman C., et al.* Zerocash: Decentralized anonymous payments from bitcoin // IEEE Symp. Security and Privacy. San Jose, USA, 2014. P. 459–474.
16. *Virza M.* On Deploying Succinct Zero-Knowledge Proofs. PhD Thesis. Massachusetts Institute of Technology, 2017. 131 p.
17. `http://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf` — Ethereum White Paper.
18. *Galal H. S. and Youssef A. M.* Verifable sealed-bid auction on the Ethereum blockchain // Intern. Conf. Financial Cryptography and Data Security. Springer, 2018. P. 265–278.
19. *Eberhardt J. and Tai S.* ZoKrates — scalable privacy-preserving off-chain computations // IEEE Intern. Conf. Blockchain. Halifax, Canada, 2018. P. 1084–1091.
20. `https://github.com/scipr-lab/libsnark` — libsnark: a C++ library for zkSNARK proofs.
21. `http://ethdocs.org/en/latest` — Ethereum Homestead Documentation.
22. `https://www.jsonrpc.org/specification` — JSON-RPC 2.0 Specification.
23. *Svensson C.* Blockchain: Using cryptocurrency with Java // Java Magazine. 2017. January/February. P. 36–46.
24. `https://docs.web3j.io/index.html` — Web3j documentation.
25. `https://solidity.readthedocs.io/en/v0.4.24` — Solidity documentation.
26. *Wood G.* Ethereum: A Secure Decentralised Generalised Transaction Ledger. `http://gavwood.com/Paper.pdf`
27. *Кондырев Д. О.* Разработка системы проведения тендеров на основе платформы Ethereum // Материалы 55-й Междунар. научн. студ. конф. МНСК-2017. Информационные технологии. Новосибирск, Новосибирский государственный университет, 2017. С. 53.

REFERENCES

1. *Wattenhofer R.* The Science of the Blockchain. 1st ed. Inverted Forest Publishing, 2016. 115 p.
2. *Nakamoto S.* Bitcoin: A Peer-to-Peer Electronic Cash System. `https://bitcoin.org/bitcoin.pdf`
3. *Raval S.* Decentralized Applications. Harnessing Bitcoin's Blockchain Technology. O'Reilly, 2016. 118 p.
4. *Kuznetsov K. V.* Konkurentnye zakupki: torgi, tendery, konkursy [Competitive Procurement: Bidding, Tendering, Contests]. St. Petersburg, Piter Publ., 2005. 368 p. (in Russian)
5. *Kondyrev D. O., Bobrov V. S., Efremov I. E., and Vlasov V. N.* Sistema provedeniya tenderov na osnove platformy Ethereum [Ethereum-Based Tender System]. Vestnik NSU, Ser. Information Technologies, 2017, vol. 15, no. 3, pp. 31–39. (in Russian)
6. *Hardwick F. S., Akram R. N., and Markantonakis K.* Fair and transparent blockchain based tendering framework — A step towards open governance. IEEE Intern. Conf. TrustCom/BigDataSE, New York, USA, 2018, pp. 1342–1347.
7. *Heilman E., Baldimtsi F., and Goldberg S.* Blindly signed contracts: Anonymous on-blockchain and off-blockchain bitcoin transactions. Intern. Conf. Financial Cryptography and Data Security, Springer, 2016, pp. 43–60.
8. *Ruffing T., Moreno-Sanchez P., and Kate A.* Coinshuffle: practical decentralized coin mixing for bitcoin. ESORICS 2014, LNCS, 2014, vol. 8713, pp. 345–364.
9. *Bissias G., Ozisik A. P., Levine B. N., and Liberatore M.* Sybil-resistant mixingfor bitcoin. Proc. WPES'14, Scottsdale, Arizona, USA, November 2014, pp. 149–158.
10. *Bonneau J., Narayanan A., Miller A., et al.* Mixcoin: anonymity for bitcoin with accountable mixes. Intern. Conf. Financial Cryptography and Data Security, Springer, 2014, pp. 486–504.

11. *Schneier B.* Applied Cryptography. Protocols, Algorithms, and Source Code in C. John Wiley & Sons, 1996. 784 p.
12. *Goldwasser S., Micali S., and Rackof C.* The knowledge complexity of interactive proof systems. STOC'85. Proc. 17th Ann. ACM Symp. Theory of Computing, Providence, Rhode Island, USA, 1985, pp. 291–304.
13. *Blum M., Feldman P., and Micali S.* Non-interactive zero-knowledge proof systems and applications. STOC'88. Proc. 20th Ann. ACM Symp. Theory of Computing, Chicago, USA, 1988, pp. 103–112.
14. *Ben-Sasson E., Chiesa A., Genkin D., et al.* SNARKs for C: Verifying program executions succinctly and in zero knowledge. CRYPTO'2013, LNCS, 2013, vol. 8043, pp. 90–108.
15. *Ben-Sasson E., Chiesa A., Garman C., et al.* Zerocash: Decentralized anonymous payments from bitcoin. IEEE Symp. Security and Privacy, San Jose, USA, 2014, pp. 459–474.
16. *Virza M.* On Deploying Succinct Zero-Knowledge Proofs. PhD Thesis. Massachusetts Institute of Technology, 2017. 131 p.
17. `http://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf` — Ethereum White Paper.
18. *Galal H. S. and Youssef A. M.* Verifable sealed-bid auction on the Ethereum blockchain. Intern. Conf. Financial Cryptography and Data Security, Springer, 2018, pp. 265–278.
19. *Eberhardt J. and Tai S.* ZoKrates — scalable privacy-preserving off-chain computations. IEEE Intern. Conf. Blockchain, Halifax, Canada, 2018, pp. 1084–1091.
20. `https://github.com/scipr-lab/libsnark` — libsnark: a C++ library for zkSNARK proofs.
21. `http://ethdocs.org/en/latest` — Ethereum Homestead Documentation.
22. `https://www.jsonrpc.org/specification` — JSON-RPC 2.0 Specification.
23. *Svensson C.* Blockchain: Using cryptocurrency with Java. Java Magazine, 2017, January/February, pp. 36–46.
24. `https://docs.web3j.io/index.html` — Web3j documentation.
25. `https://solidity.readthedocs.io/en/v0.4.24` — Solidity documentation.
26. *Wood G.* Ethereum: A Secure Decentralised Generalised Transaction Ledger. `http://gavwood.com/Paper.pdf`
27. *Kondyrev D. O.* Razrabotka sistemy provedeniya tenderov na osnove platformy Ethereum [Development of a tender system based on the Ethereum platform]. Proc. MNSK-2017, Information Technology, Novosibirsk, NSU Publ., 2017, p. 53. (in Russian)

CRYPTOLOGIA
2020, VOL. 44, NO. 3, 223–256
https://doi.org/10.1080/01611194.2019.1670282

Taylor & Francis
Taylor & Francis Group

The Fifth International Students' Olympiad in cryptography—NSUCRYPTO: Problems and their solutions

Anastasiya Gorodilova, Sergey Agievich, Claude Carlet, Xiang-dong Hou, Valeria Idrisova, Nikolay Kolomeec, Alexandr Kutsenko, Luca Mariot, Alexey Oblaukhov, Stjepan Picek, Bart Preneel, Razvan Rosie, and Natalia Tokareva

ABSTRACT
Problems and their solutions of the Fifth International Students' Olympiad in cryptography NSUCRYPTO'2018 are presented. We consider problems related to attacks on ciphers and hash functions, Boolean functions, quantum circuits, Enigma, etc. We discuss several open problems on orthogonal arrays, Sylvester matrices, and disjunct matrices. The problem of existing an invertible Sylvester matrix whose inverse is again a Sylvester matrix was completely solved during the Olympiad.

KEYWORDS
hash functions; Enigma; quantum circuits; metrically regular sets; irreducible polynomials; orthogonal arrays; Sylvester matrices; disjunct matrices; Olympiad; NSUCRYPTO

Introduction

NSUCRYPTO—The International Students' Olympiad in cryptography—celebrated its 5-year anniversary in 2018. Interest in the Olympiad around the world is significant: there were more than 1,600 participants from 52 countries in the first five Olympiads from 2014 to 2018! The Olympiad program committee includes specialists from Belgium, France, The Netherlands, USA, Norway, India, Belarus', and Russia.

Let us shortly formulate the format of the Olympiad. One of the Olympiad main ideas is that everyone can participate! Each participant chooses his/her category when registering on the Olympiad website nsucrypto.nsu.ru. There are three categories: "school students" (for junior researchers: pupils and high school students), "university students" (for participants who are currently studying at universities), and "professionals" (for participants who have already completed education or just want to be in the restriction-free category). Awarding of the winners is held in each category separately.

The Olympiad consists of two independent Internet rounds: the first one is individual (duration 4 hours 30 minutes) while the second round is team (duration 1 week). The first round is divided into two sections: A—for

CONTACT A. Gorodilova gorodilova@math.nsc.ru Novosibirsk State University, Novosibirsk, 630090 Russia; Sobolev Institute of Mathematics, Novosibirsk, 630090 Russia.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ucry.

"school students," B—for "university students" and "professionals." The second round is general for all participants. Participants read the Olympiad problems and submit their solutions using the Olympiad website. The language of the Olympiad is English.

The Olympiad participants are always interested in solving different problems of various complexities at the intersection of mathematics and cryptography. They show their knowledge, creativity, and professionalism. That is why the Olympiad not only includes interesting tasks with known solutions but also offers unsolved problems in this area. This year, one of such open problems, "Sylvester matrices," was completely solved by three teams! All the open problems stated during the Olympiad history can be found at nsucrypto.nsu.ru/unsolved-problems. On the website we also mark the current status of each problem. For example, in addition to "Sylvester matrices" solved in 2018, the problem "algebraic immunity" was completely solved during the Olympiad in 2016. And what is important for us, some participants were trying to find solutions after the Olympiad was over. For example, a partial solution for the problem "A secret sharing" (2014) was proposed in Geut et al. (2017). We invite everybody who has ideas on how to solve the problems to send your solutions to us!

The paper is organized as follows. We start with problem structure of the Olympiad in each section (Problem structure of the Olympiad). Then we present formulations of all the problems stated during the Olympiad and give their detailed solutions in each section (Problems and their solutions). Finally, we publish the lists of NSUCRYPTO'2018 winners in each section (Winners of the Olympiad).

Mathematical problems of the previous International Olympiads NSUCRYPTO'2014, NSUCRYPTO'2015, NSUCRYPTO'2016, and NSUCRYPTO'2017 can be found in Agievich et al. (2015, 2017), Tokareva et al. (2018), and Gorodilova et al. (2019), respectively.

Problem structure of the Olympiad

There were 16 problems stated during the Olympiad, and some of them were included in both rounds (Tables 1 and 2). Section A of the first round consisted of six problems, whereas section B contained seven problems. Three problems were common for both sections. The second round was composed of 11 problems. Three problems of the second round were marked as unsolved (awarded special prizes from the Program Committee).

Problems and their solutions

In this section we formulate all the problems of NSUCRYPTO'2018 and present their detailed solutions paying attention to solutions proposed by the participants.

Table 1. Problems of the first round.

Section A

N	Problem title	Maximum scores
1	A digital signature	4
2	Jack and the Beanstalk	4
3	Key matrices	4
4	A sequence	4
5	Solutions of the equation	4
6	Stickers	6

Section B

N	Problem title	Maximum scores
1	Stickers	6
2	Key matrices	4
3	A sequence	4
4	Quantum circuits	4
5	Bash-S3	8
6	Metrical cryptosystem—2	6
7	A fixed element	10

Table 2. Problems of the second round.

N	Problem title	Maximum scores
1	A digital signature	4
2	Orthogonal arrays	Unsolved
3	Hash function `FNV-1a`	8
4	`TwinPeaks2`	6
5	An Enigmatic Challenge	8
6	Sylvester matrices	Unsolved
7	Stickers	6
8	`Bash-S3`	8
9	Metrical cryptosystem—2	6
10	A fixed element	10
11	Disjunct Matrices	Unsolved

Problem "A digital signature"

Formulation

Alice uses a new digital signature algorithm, that turns a text message M into a pair (M, s), where s is an integer and generated in the following way:

1. The special function h transforms M into a big positive integer $r = h(M)$.
2. The number $t = r^2$ is calculated, where $t = \overline{t_1 t_2 \dots t_n}$.
3. The signature s is calculated as $s = t_1 + t_2 + \cdots + t_n$.

Bob obtained the signed message

(Congratulations on the fifth year anniversary of NSUCRYPTO!, 2018)

from Alice and immediately recognized that something was wrong with the signature! How did he discover it?

Remarks. By $t = \overline{t_1 t_2 \dots t_n}$ we mean that $t_1, t_2, \dots, t_n$ are decimal digits and all digits under the bar form decimal number t.

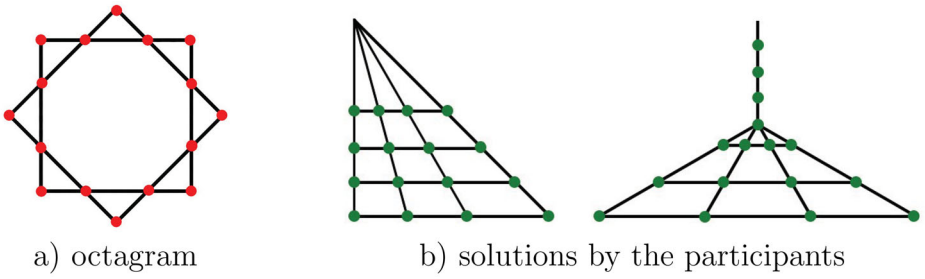


Figure 1. Lines and seeds.

Solution

It is widely known that every integer is congruent to the sum of its digits modulo 3. So, we have that $t \equiv_3 2018 \equiv_3 2$. But t is equal to r^2 and a square can not be equal to 2 modulo 3. Thus, we have a contradiction.

We got a lot correct solutions. The most accurate and detailed solutions were sent by Ruxandra Icleanu (Tudor Vianu National College of Computer Science, Romania), Petr Ionov (Yaroslavl State University, Russia), and the team of Henning Seidler and Katja Stumpp (TU Berlin, Germany).

Problem "Jack and the Beanstalk"

Formulation

Little Jack is only 7 years old and likes solving riddles involving the powers of two. Recently, his uncle Bitoshi gave him 16 BeanCoin seeds and promised that Jack can collect all BeanCoins which will grow from these seeds. But in order for BeanCoins to grow big and fruitful, Jack must plant the seeds in the garden in a special way. He has to draw eight lines on the ground and plant all 16 seeds on these lines in such a way that each of the lines contains exactly four seeds.

Can you help Jack to achieve his goal and suggest how to plant the seeds?

Solution

The seeds can be placed on the corners and intersection points of an octagram, as depicted in Figure 1a. As is clear from this figure, all eight lines contain exactly four seeds and it is impossible to draw other line contained exactly four seeds.

Many school students found interesting ways to draw these lines, for example Figure 1b. The most interesting ones were given by Gorazd Dimitrov (Yahya Kemal College, Macedonia), Artem Ismagilov (The Specialized Educational and Scientific Center UrFU, Russia), and Igor Pastushenko (The Specialized Educational Scientific Center of Novosibirsk State University, Russia).

Problem "Key matrices"

Formulation

Let n be an odd positive integer. In some cipher, a key is a binary $n \times n$ matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix},$$

where $a_{i,j}$ is either 0 or 1, such that each diagonal of any length $1, 2, \dots, n-1, n$ contains an **odd** number of 1s.

What is the minimal and the maximal number of 1s that can be placed in a key matrix A?

Remarks. For example, for $n = 3$, diagonals are the following ten lines:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$$

Solution

The correct solution of this problem must consist of two steps. The first step is to find theoretical lower and upper bounds for the number of 1s, and the second step is to prove that these bounds are tight. The best solution was proposed by Aleksei Udovenko (University of Luxembourg), which we provide below.

1. Minimum. Consider the $n \times n$ matrix A (n is odd) with both the top row filled with 1s, the bottom row filled with 1s, and the central cell equal to 1; all other elements are 0:

$$\begin{cases} a_{1,i} = 1, & 1 \le i \le n; \\ a_{n,i} = 1, & 1 \le i \le n; \\ a_{(n+1)/2,(n+1)/2} = 1; & \\ a_{i,j} = 0, & \text{otherwise.} \end{cases}$$

Any diagonal of length less than $n - 1$ includes exactly a single 1 (either from the top row or from the bottom row). The two diagonals of length n include three 1s (one from the top row, one from the bottom row, and one from the center). Therefore, this matrix satisfies the condition. It has $2n + 1$ 1s.

We now prove that this number of 1s is minimal. Note that each corner cell $a_{1,1}, a_{1,n}, a_{n,1}, a_{n,n}$ makes a single element diagonal. Therefore, these cells must contain 1s. There are $2(n-2)$ diagonals going in the down-right direction and

not touching the corners (starting from the cells of the leftmost column and from the cells for the topmost row). Furthermore, the main diagonal without the corner cells must have an odd number of 1s too. Therefore, $2n-3$ disjoint diagonals must contain at least one 1, in addition to 4 corner 1s. Therefore, there should be at least $2(n-2)+1+4=2n+1$ 1s in the matrix.

2. Maximum. Consider the $n \times n$ matrix A (n is odd) filled with 1s except cells in the leftmost and the rightmost columns which have an even row index:

$$\begin{cases} a_{2i,1}=0, & 1 \le i \le (n-1)/2; \\ a_{2i,n}=0, & 1 \le i \le (n-1)/2; \\ a_{i,j}=1, & \text{otherwise.} \end{cases}$$

It is easy to check that all diagonals that contain an even number of elements contain a single zero either from the leftmost or from the rightmost column. Therefore, these diagonals have an odd number of 1s. Also, all diagonals that contain an odd number of elements contain no zeroes and thus have an odd number of 1s too. Therefore, this matrix satisfies the condition. It has $n^2-2(n-1)/2=n^2-n+1$ 1s.

We now prove that this number is maximal. Consider diagonals going in the down-right direction that have an even number of elements. There are $2(n-1)/2=(n-1)$ such diagonals and they are disjoint. Each of them must contain at least a single zero. Therefore, the maximum number of 1s is n^2-n+1.

Problem "A sequence"

Formulation

Two friends, Roman and Anton, are very interested in sequences and ciphers. Their new cryptosystem encrypts binary messages of length n, $X=(x_1,x_2,\ldots,x_n)$, where each x_i is either 0 or 1. A key K of the cipher is a set of n integers $a_1,a_2,\ldots,a_n$. The ciphertext Y for the message X encrypted with the key K is the integer

$$Y = x_1 \cdot a_1 + x_2 \cdot a_2 + \cdots + x_n \cdot a_n.$$

Roman and Anton change their key regularly. Today, the key K is defined by

$$a_i = 2^i + (-1)^i \text{ for all } i = 1,\ldots,n.$$

The friends can easily decipher any message using the key defined by this sequence for any n!

1. Prove that the encryption is correct for this key K for any n: there are no two distinct input messages X^1 and X^2 such that their ciphertexts Y^1 and Y^2 are equal, i.e., $Y^1 = Y^2$.

2. Describe an algorithm which can be used to easily decipher any ciphertext Y encrypted with today's key K. Here "easily" means that the algorithm should work much faster than checking all possible variants for an input message X.

Solution

Let us firstly show that the sequence $\{a_i\}$ is *superincreasing*, i.e., $a_{i+1} > \sum_{k=1}^{i} a_k$ for any $i > 0$. Indeed,

$$\sum_{k=1}^{i} a_k = \sum_{k=1}^{i} (2^k + (-1)^k) = 2^{i+1} - 2 + \sum_{k=1}^{i} (-1)^k$$
$$= \begin{cases} 2^{i+1} - 2, & \text{if } i \text{ is even} \\ 2^{i+1} - 3, & \text{if } i \text{ is odd} \end{cases} < 2^{i+1} + (-1)^i = a_{i+1}.$$

1. Let us show that the encryption is correct. Let $X^1 = (x_1^1, \ldots, x_n^1)$ and $X^2 = (x_1^2, \ldots, x_n^2)$ be two distinct messages, and i is the largest position such that $x_i^1 \neq x_i^2$. Without loss of generality, suppose that $x_i^1 = 1$. Then

$$Y^1 - Y^2 = (x_1^1 \cdot a_1 + \cdots + x_i^1 \cdot a_i + \cdots + x_n^1 \cdot a_n)$$
$$-(x_1^2 \cdot a_1 + \cdots + x_i^2 \cdot a_i + \cdots + x_n^2 \cdot a_n)$$
$$= (x_1^1 - x_1^2) \cdot a_1 + \cdots + (x_{i-1}^1 - x_{i-1}^2) \cdot a_{i-1} + a_i > 0$$

since $\{a_i\}$ is a superincreasing sequence.

2. The correctness of the decryption algorithm (Algorithm 1) is also based on the superincreasing property of $\{a_i\}$. The complexity of the algorithm consists of n integer comparisons.

Algorithm 1. The decryption algorithm

Input: Y, n.
Output: $X = (x_1, \ldots, x_n)$.
Step 0. $T := Y, i := n$.
Step 1. If $T > a_i$, then $x_i = 1$; else $x_i = 0$.
Step 2. $T := T - x_i \cdot a_i, i := i - 1$. If $i > 0$, go to step 1; else return X.

The problem was solved by the majority of participants including eight school students.

Problem "Solutions of the equation"

Formulation

Alice is studying special functions that are used in symmetric ciphers. Let E^n be the set of all binary vectors $x = (x_1, x_2, ..., x_n)$ of length n, where x_i is either 0 or 1. Given two vectors x and y from E^n consider their sum $x \oplus y = (x_1 \oplus y_1, ..., x_n \oplus y_n)$, where $\oplus$ is addition modulo 2.

Example. If $n = 3$, then $E^3 = \{(000), (001), (010), (011), (100), (101), (110), (111)\}$. Let $x = (010)$ and $y = (011)$, then vector $x \oplus y$ is equal to $(010) \oplus (011) = (0 \oplus 0, 1 \oplus 1, 0 \oplus 1) = (001)$.

We will say that a function F maps E^n to E^n if it transforms any vector x from E^n into some vector $F(x)$ from E^n.

Example. Let $n = 2$. For instance, we can define F that maps E^2 to E^2 as follows: $F(00) = (00), F(01) = (10), F(10) = (11)$ and $F(11) = (10)$.

Alice found a function S that maps E^6 to E^6 in such a way that the vectors $S(x)$ and $S(y)$ are not equal for any nonequal vectors x and y. Also, S has another curious property: the equation

$$S(x) \oplus S(x \oplus a) = b$$

has either 0 or 2 solutions for any nonzero vector a from E^6 and any vector b from E^6.

Find the number of pairs (a, b) such that this equation has exactly two solutions!

Solution

Consider a function S that satisfies the conditions of the problem. Let us fix an arbitrary vector a that is nonzero. Consider the set B_a of all possible values of $S(x) \oplus S(x \oplus a)$, i.e., $B_a = \{S(x) \oplus S(x \oplus a) \mid x \in E^6\}$. It holds that $|B_a| = 2^5$, since $S(x) \oplus S(x \oplus a) = S(x \oplus a) \oplus S(x \oplus a \oplus a)$. Then for every nonzero a there exist 2^5 values of b, such that $S(x) \oplus S(x \oplus a) = b$ has two solutions. Then the number of pairs is equal to $63 * 32 = 2016$.

Correct answers were sent by only three school students: Alexey Lvov (Gymnasium 6 of Novosibirsk, Russia), Borislav Kirilov (The First Private Mathematical Gymnasium of Sofia, Bulgaria), and Razvan Andrei Draghici (National College Fratii Buzesti, Romania).

Problem "Quantum circuits"

Formulation

Alice and Bob are interested in quantum circuits. They studied quantum operations and would like to use them for their simple cipher. Let an input plaintext

Table 3. Quantum gates and circuit symbols.

Pauli-X gate	$\|x\rangle$ — X — $\|x \oplus 1\rangle$	Acts on a single qubit in the state $\|x\rangle, x \in \{0,1\}$.
Controlled-NOT gate (CNOT gate)	$\|x\rangle$ — $\|x\rangle$ $\|y\rangle$ — $\|y \oplus x\rangle$	Acts on two qubits in the states $\|x\rangle, \|y\rangle$, $x, y \in \{0,1\}$; it flips the second qubit if and only if the first qubit is in the state $\|1\rangle$.
Toffoli gate (CCNOT gate)	$\|x\rangle$ — $\|x\rangle$ $\|y\rangle$ — $\|y\rangle$ $\|z\rangle$ — $\|z \oplus (x \wedge y)\rangle$	Acts on three qubits in the states $\|x\rangle, \|y\rangle, \|z\rangle, x, y, z \in \{0,1\}$; it flips the third qubit if and only if the states of the first and the second qubits are both equal to $\|1\rangle$.
$\|x\rangle$ — x		A measurement of a qubit in the state $\|x\rangle, x \in \{0,1\}$, in the computational basis $\{\|0\rangle, \|1\rangle\}$.
		A wire carrying a single qubit (time goes left to right).
		A wire carrying a single classical bit.

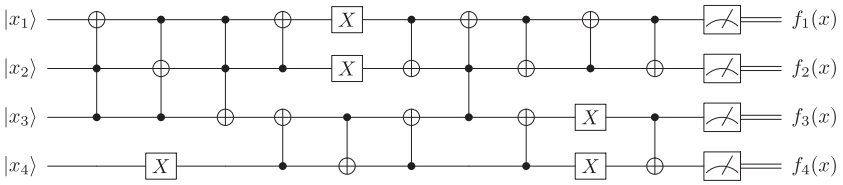
be $P = (p_1, p_2, ..., p_{16}) \in \mathbb{F}_2^{16}$. The ciphertext $C \in \mathbb{F}_2^{16}$ is calculated as

$$C = K \oplus (F(p_1, ..., p_4), F(p_5, ..., p_8), F(p_9, ..., p_{12}), F(p_{13}, ..., p_{16})),$$

where $K \in \mathbb{F}_2^{16}$ is a secret key and F is a function from $\mathbb{F}_2^4$ to $\mathbb{F}_2^4$; $\oplus$ is bit-wise XOR.

The friends found a representation of F from wires and elementary quantum gates which form a quantum circuit. They use Dirac notation and denote computational basis states by $|0\rangle$ and $|1\rangle$. Further, quantum bits (qubits) are considered only in quantum states $|0\rangle$ and $|1\rangle$. Alice and Bob used the following quantum gates and circuit symbols which are given in Table 3.

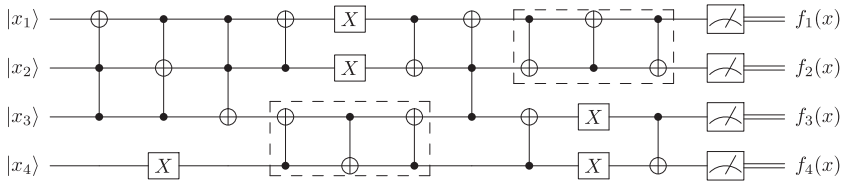
A quantum circuit which describes action of F on $x = (x_1, x_2, x_3, x_4) \in \mathbb{F}_2^4$, where $F = (f_1, f_2, f_3, f_4)$ and $f_i, i = 1, 2, 3, 4$, are Boolean functions in four variables, is the following:



The problem. The friends encrypted the plaintext $P = (0011010111110010)$ and got the ciphertext $C = (1001101010010010)$. Find the secret key K!

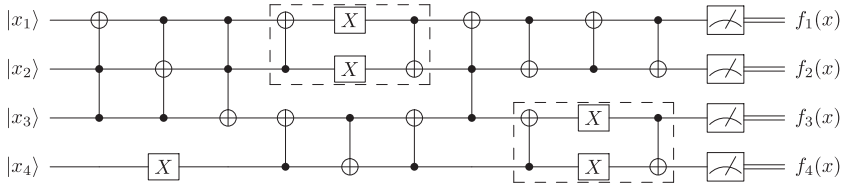
Solution

One can notice that the given circuit can be simplified by observing that the following evolutions

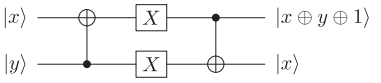


actually swap two states $|x\rangle, |y\rangle, x, y \in \{0, 1\}$:

Both of the evolutions



have form



for $|x\rangle, |y\rangle, x, y \in \{0, 1\}$.

The algebraic normal forms of coordinate Boolean functions of F are

$$\begin{aligned} f_1(x) &= x_1 \oplus x_2x_3, \\ f_2(x) &= x_2 \oplus x_1x_4 \oplus x_2x_3x_4 \oplus 1, \\ f_3(x) &= x_3 \oplus x_4 \oplus x_1x_2 \oplus x_1x_3, \\ f_4(x) &= x_4 \oplus 1, \end{aligned}$$

where $x \in \mathbb{F}_2^4$. Then

$$\begin{aligned} K_{1,\dots,4} &= C_{1,\dots,4} \oplus F(p_1, \dots, p_4) = C_{1,\dots,4} \oplus (0100) = (1101), \\ K_{5,\dots,8} &= C_{5,\dots,8} \oplus F(p_5, \dots, p_8) = C_{5,\dots,8} \oplus (0010) = (1000), \\ K_{9,\dots,12} &= C_{9,\dots,12} \oplus F(p_9, \dots, p_{12}) = C_{9,\dots,12} \oplus (0000) = (1001), \\ K_{13,\dots,16} &= C_{13,\dots,16} \oplus F(p_{13}, \dots, p_{16}) = C_{13,\dots,16} \oplus (0111) = (0101), \end{aligned}$$

and finally, the key is the following:

$$K = (1101100010010101).$$

Many participants coped with this problem and correctly found the key.

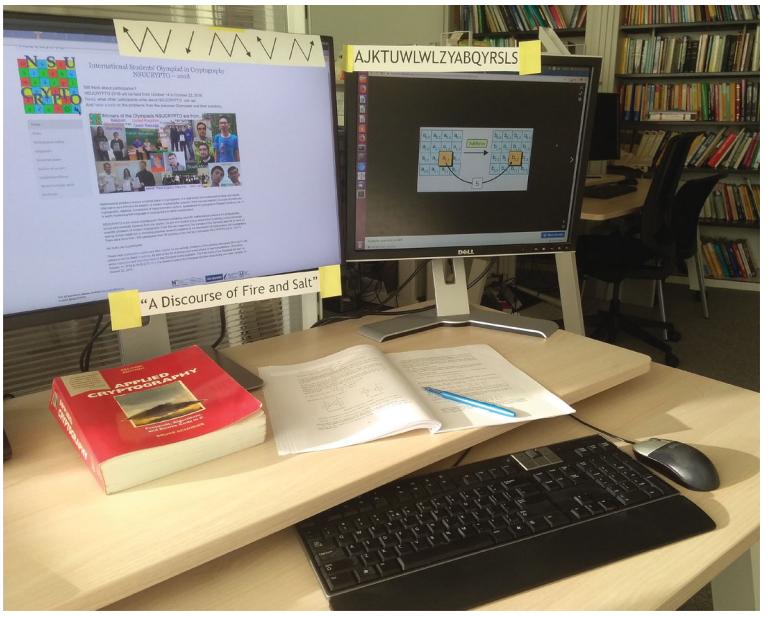


Figure 2. Workspace.

Problem "Stickers"

Formulation

Bob always takes into account all the recommendations of security experts. He switched from short passwords to long passphrases and changes them every month. Bob usually chooses passphrases from the books he is reading. Passphrases are so lengthy and are changed so often! In order to not forget them, Bob decided to use stickers with hints. He places them on his monitors (ooh, experts …). The only hope is that Bob's hint system is reliable because it uses encryption. But is that true? Could you recover Bob's current passphrase from the photo of his workspace (Figure 2)?

Solution

Looking at the picture we see three stickers. One of them is "A Discourse of Fire and Salt" that represents a title of a book written by Blaise de Vigenère. This is the first hint that probably the Viginère cipher was used. Then we have a sticker with the ciphertext `AJKTUWLWLZYABQYRSLS` that consists of 19 letters. And finally, we see the sticker with five directed polygonal paths containing a total of 19 vertices. These 19 vertices could correspond to the 19 ciphertext letters.

There is a keyboard at the picture. So, we can guess that these arrows could be related to the letters from the keyboard. Let us look at the first two keyboard rows (Figure 3). We can recover the secret key

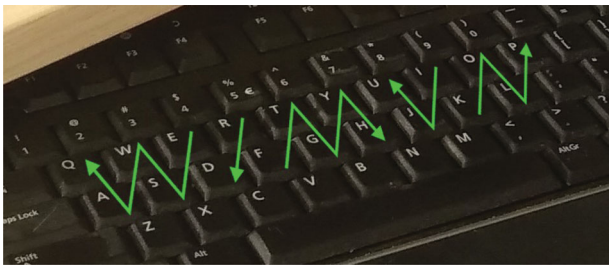


Figure 3. Keyboard rows.

ESWAQRDFTGYHIJUKOLP. By deciphering the ciphertext using this key and the Viginère cipher, we get WROTEFIRSTATTHEHEAD. Thus, Bob's current passphrase is "Wrote first at the head."

Surprisingly, nobody solved this problem in the first round, while five teams solved it in the second round.

Problem "Bash-S3"

Formulation

The sponge function Bash-f (Agievich et al. 2016) uses the permutation *S3* that transforms a triple of 64-bit binary words a, b, c in the following way:

$$S3(a,b,c) = (b \vee \neg c \oplus a, a \vee c \oplus b, a \wedge b \oplus c).$$

Here $\neg, \wedge, \vee, \oplus$ denote the binary bitwise operations "NOT," "AND," "OR," "XOR," respectively. The operations are listed in descending order of priority. Let w^k also denote the cyclic shift of a 64-bit word w to the left by $k \in \{1, 2, \ldots, 63\}$ positions.

Alice wants to strengthen S3. She can do this by XORing any input a, b, c or its cyclic shift to any output. She must use at least one cyclic shift and she cannot add two identical terms to the same output.

Help Alice change S3 in such a way that a modified S3 will still be a permutation!

Remarks.

1. For example, in the expression $b \vee \neg c \oplus a$, we firstly calculate $\neg c$, then calculate $b \vee \neg c$, and after that the final result (according to descending order of operations priority).
2. The modification

$$(b \vee \neg c \oplus a \oplus a^{11}, a \vee c \oplus a^7 \oplus c, a \wedge b \oplus b^{32})$$

is allowed but it does not satisfy the permutation condition.

3. S3 has three outputs: $b \vee \neg\, c \oplus a, a \vee c \oplus b, a \wedge b \oplus c$. Alice can add as many inputs and cyclic shifts of inputs as she wants to each of these outputs. In the remark 2 she adds a^{11} to the first output, $b \oplus a^7 \oplus c$ to the second output, and $c \oplus b^{32}$ to the third output. Note that the fact that S3 is a permutation (as a function $\{0,1\}^{64*3} \to \{0,1\}^{64*3}$) is not obvious. But the problem is only to prove that the modification of S3 is a permutation too (as a function $\{0,1\}^{64*3} \to \{0,1\}^{64*3}$).

Solution

It is allowed to add to the outputs of S3 the outputs of the following linear transformation:

$$L(a,b,c) = (L_0(a,b,c), L_1(a,b,c), L_2(a,b,c))$$

that is defined by bitwise XOR operations and cyclic shifts.

The permutation property of a modified S3 will be broken if for some distinct $(a,\ b,\ c)$ and (a',b',c')

$$S3(a,b,c) \oplus S3(a',b',c') = L(a,b,c) \oplus L(a',b',c'). \tag{1}$$

We will call the expressions from both sides of equality (1) and the sum $(a,b,c) \oplus (a',b',c')$ by *differences*. Let

$$(w_0, w_1, w_2) = (a,b,c) \oplus (a',b',c'),$$
$$(W_0, W_1, W_2) = S3(a,b,c) \oplus S3(a',b',c').$$

On the one hand, input and output differences of S3 satisfy (for instance, see Agievich et al. 2016) the equality

$$w_0 \wedge W_0 \oplus w_1 \wedge W_1 \oplus w_2 \wedge W_2 = 11...1.$$

On the other hand, by (1) the permutation property of a modified S3 will be broken if

$$(W_0, W_1, W_2) = L(a,b,c) \oplus L(a',b',c') = L(w_0, w_1, w_2).$$

As a result, a modified S3 will be still a permutation if the following equality

$$w_0 \wedge L_0(w_0,w_1,w_2) \oplus w_1 \wedge L_1(w_0,w_1,w_2) \oplus w_2 \wedge L_2(w_0,w_1,w_2) = 11...1 \tag{2}$$

does not hold for any nonzero (w_0, w_1, w_2). For example, if

$$L(a,b,c) = (a \oplus a^d \oplus b, a \oplus c, b), \quad d \in \{1,2,...,63\},$$

then (2) becomes

$$w_0 \wedge (w_0 \oplus w_0^d \oplus w_1) \oplus w_1 \wedge (w_0 \oplus w_2) \oplus w_2 \wedge (w_1) = w_0 \wedge (w_0 \oplus w_0^d) \neq 11...1.$$

Thus, we found the following solution for the problem:

$$S3(a,b,c) \oplus L(a,b,c) = (b \vee \neg\, c \oplus a^d \oplus b, a \vee c \oplus a \oplus b \oplus c, a \wedge b \oplus b \oplus c).$$

Note that there are many other possible solutions.

This problem was completely solved by three participants in the first round and by nine teams in the second round. Many of these solutions were interesting and compact.

Problem "Metrical cryptosystem—2"

Formulation

Let $\mathbb{F}_2^n$ be an n-dimensional vector space over the field $\mathbb{F}_2 = \{0, 1\}$. Alice and Bob exchange messages using the following cryptosystem.

1. First, they use a supercomputer to calculate two special large secret sets $A, B \subseteq \mathbb{F}_2^n$ which have the following property: there exists a constant ℓ ($\ell \geq 26$), such that for any $x \in \mathbb{F}_2^n$ it holds
$$d(x, A) + d(x, B) = \ell,$$
where $d(x, A)$ denotes Hamming distance from the vector x to the set A.
2. Alice then saves the number ℓ, the set A and a set of vectors $a_1, a_2, \ldots, a_r$ such that for any $k : 0 \leq k \leq \ell$, there is a vector a_{i} at distance k from A. Similarly, Bob saves the number ℓ, the set B and a set of vectors $b_1, b_2, \ldots, b_s$ such that for any $k : 0 \leq k \leq \ell$, there is a vector b_{i} at distance k from B.
3. Text messages are encrypted letter by letter. In order to encrypt a letter Alice replaces it with its number in the alphabet, say k. Then she chooses some vector a_i at distance k from the set A and sends this vector over to Bob. Bob then calculates the distance $d(a_i, B)$ and using the property of the sets A, B, calculates $k = \ell - d(a_i, B)$. So, he gets the letter Alice sent. If Bob wants to send an encrypted message to Alice, he does the same but using his saved vectors and the set B.

Eve was able to hack the supercomputer when it was calculating the sets A and B. She extracted the set C from its memory, which consists of all vectors of $\mathbb{F}_2^n$ that are at distance 1 or less from either A or B. She also learned that ℓ is even.

Help Eve to crack the presented cryptosystem (to decrypt any short intercepted message)! You know that she has (illegal) access to the supercomputer, which can calculate and output the list of distances from all vectors of $\mathbb{F}_2^n$ to any input set D in reasonable (but not negligible) time.

Remarks. Recall several definitions and notions. The *Hamming distance* $d(x, y)$ between vectors x and y is the number of coordinates in which these vectors differ. Distance from vector $y \in \mathbb{F}_2^n$ to the set $X \subseteq \mathbb{F}_2^n$ is defined as $d(y, X) = \min_{x \in X} d(y, x)$.

Solution

Let us denote by A_i (B_i respectively) the set of all vectors at distance i from the set A (B respectively):

$$A_i = \{x \in \mathbb{F}_2^n : d(x, A) = i\}, B_i = \{x \in \mathbb{F}_2^n : d(x, B) = i\}.$$

It is easy to see that

- $A = A_0 = B_\ell$,
- $B = B_0 = A_\ell$,
- $A_i = B_{\ell - i}$ for any $i \in \{0, \ldots, \ell\}$,
- $C = A_1 \cup B_1 \cup A_0 \cup B_0$.

From the definition of the Hamming distance it is easy to prove that if a vector x lies in the set A_i, then it is at distance $|i-j|$ from the set A_j for any i, j.

Proof . Indeed, if $i = j$, the statement is trivial.

Assume that $i > j$. By definition, $d(x, A) = i$, so there exists a shortest path of length i from A to x, consisting of vectors $x_0, x_1, \ldots, x_i = x$, where $x_0 \in A$. Since consecutive vectors in the path differ in only one coordinate, and vectors from A_s and A_t can be neighbors only if $|s-t| \leq 1$, it follows that $x_k \in A_k$ for every $k = 0, \ldots, i$. So, vector x_j from the path belongs to A_j and is at distance $i - j$ from vector x. Therefore, $d(x, A_j) \leq i-j$. Distance cannot be less than $i - j$, because then $d(x, A)$ would have been less than i, which contradicts conditions of the statement. Thus, $d(x, A_j) = i-j$.

If $i < j$, then we can replace A_i with $B_{\ell-i}$, A_j with $B_{\ell-j}$ and use B instead of A for the same argument as in the previous case. □

In particular, given x is in A_i, it is at distance $|i-1|$ from the set A_1 and at distance $|i-(\ell-1)|$ from the set B_1.

Let us "feed" the set C to the supercomputer. We denote the maximal distance from vectors of $\mathbb{F}_2^n$ to vectors of C as r, and the set of all vectors achieving this distance as $\hat{C}$. Taking into account the statement proven above (and the fact that ℓ is even), we can see that the maximum is achieved for vectors of the set $A_{\frac{\ell}{2}}$. Hence, $r = \frac{\ell}{2} - 1$ and $\hat{C} = A_{\frac{\ell}{2}}$. Thus, we can calculate ℓ as $2r + 2$.

Assume now that Alice sends a message $a_{i_1}, a_{i_2}, \ldots, a_{i_k}$ to Bob. Eve intercepts it and (using the obtained table of distances from the set C) calculates

that these vectors are at distances $s_1, s_2, ..., s_k$ from the set C. Therefore, they are at distances $s_1+1, s_2+1, ..., s_k+1$ from the set $A \cup B$. Since $d(x, A \cup B) = \min(d(x, A), d(x, B))$, each encrypted letter could be either s_i+1 or $\ell-(s_i+1)$. If one of these two numbers is greater than 26, we can easily determine the encrypted letter, if not, we can consider both possibilities. In the worst case we would need to consider 2^N variants, where N is the length of the message, but since messages are short and are written in natural language, we do not need to check all of them and the decryption should not be hard.

Note: Sets A and B satisfying condition from Step 1 of the problem (for an arbitrary constant ℓ not necessarily greater than 26) are called *strongly metrically regular* and are studied in Oblaukhov (2019).

The best solutions to the problem were submitted by Alexey Chilikov (Bauman Moscow State Technical University, Russia) and Saveliy Skresanov (Novosibirsk State University, Russia).

Problem "A fixed element"

Formulation

A polynomial $f(X_1, ..., X_n) \in \mathbb{F}_2[X_1, ..., X_n]$ is called *reduced* if the degree of each X_i in f is at most 1. For $0 \le r \le n$, the rth order Reed—Muller code of length 2^n, denoted by $R(r, n)$, is the $\mathbb{F}_2$-space of all reduced polynomials in $X_1, ..., X_n$ of total degree less than or equal to r. We also define $R(-1, n) = \{0\}$.

The general linear group $\mathrm{GL}(n, \mathbb{F}_2)$ acts on $R(r, n)$ naturally: Given $A \in \mathrm{GL}(n, \mathbb{F}_2)$ and $f(X_1, ..., X_n) \in R(r, n)$, Af is defined to be the reduced polynomial obtained from $f((X_1, ..., X_n)A)$ by replacing each power X_i^k $(k \ge 2)$ with X_i. Consequently, $\mathrm{GL}(n, \mathbb{F}_2)$ acts on the quotient space $R(r, n)/R(r-1, n)$.

Let $A \in \mathrm{GL}(n, \mathbb{F}_2)$ be such that its characteristic polynomial is a primitive irreducible polynomial over $\mathbb{F}_2$. Prove that the only element in $R(r, n)/R(r-1, n)$, where $0 < r < n$, fixed by the action of A is 0.

Solution

Let $\binom{\{1, ..., n\}}{r}$ denote the set of r-subsets of $\{1, ..., n\}$. When A acts on $R(r, n)/R(r-1, n)$, its matrix with respect to the basis $\prod_{i \in I} X_i, I \in \binom{\{1, ..., n\}}{r}$, is the rth compound matrix $C_r(A)$ of A. The eigenvalues of A are $\gamma^{2^i}, 0 \le i \le n-1$, where γ is a primitive element of $\mathbb{F}_{2^n}$. The eigenvalues of $C_r(A)$ are all possible products of r eigenvalues of A, i.e.,

$$\gamma^{\sum_{i\in I} 2^i}, \quad I \in \binom{\{0,\ldots,n-1\}}{r}.$$

Clearly, the above expression never equals 1. Hence 1 is not an eigenvalue of $C_r(A)$. Therefore, the action of A does not fix any nonzero element in $R(r,n)/(r-1,n)$.

The problem was solved by four teams in the second round: Aleksei Udovenko (University of Luxembourg), the team of Dianthe Bose and Neha Rino (Chennai Mathematical Institute, India), the team of Andrey Kalachev, Danil Cherepanov and Alexey Radaev (Bauman Moscow State Technical University, Russia), and the team of Sergey Titov and Kristina Geut (Ural State University of Railway Transport, Russia).

Problem "Hash function `FNV-1a`*"*

Formulation

Hash function `FNV-1a` (http://www.isthe.com/chongo/tech/comp/fnv/) processes a message x composed of bytes $x_1, x_2, \ldots, x_n \in \{0, 1, \ldots, 255\}$ in the following way:

1. $h \leftarrow h_0$;
2. for $i = 1, 2, \ldots, n$: $h \leftarrow (h \oplus x_i)g \bmod 2^{128}$;
3. return h.

Here $h_0 = 144066263297769815596495629667062367629$, $g = 2^{88} + 315$. The expression $h \oplus x_i$ means that the least significant byte of h is added bitwise modulo 2 with the byte x_i.

Find a collision, that is, two different messages x and x' such that $\texttt{FNV}-1\texttt{a}(x) = \texttt{FNV}-1\texttt{a}(x')$. Collisions on short messages and collisions that are obtained without intensive calculations are welcomed. Supply your answer as a pair of two hexadecimal strings which encode bytes of colliding messages.

Solution

We will base the solution to the problem on "`FNV2`" (NSUCRYPTO'2017) (Gorodilova et al. 2019), where it was required to find a collision for the similar hash function `FNV2`. `FNV-1a` differs from `FNV2` in the following: instead of the $\oplus$ operation for adding h and x_i it uses standard $+$ operation.

It is easy to see that

$$\texttt{FNV2}(x_1x_2\ldots x_n) = (h_0g^n + x_1g^n + x_2g^{n-1} + \ldots + x_ng) \bmod 2^{128}.$$

For FNV2, we found a relation

$$a_1 g^{n-1} + a_2 g^{n-2} + \cdots + a_n \equiv 0 \pmod{2^{128}},$$

where $a_i \in \{-255, \ldots, 255\}$.

Then we represented a_i as the difference $x_i - x'_i$ and found a collision

$$\mathrm{FNV2}(x_1 x_2 \ldots x_n) - \mathrm{FNV2}(x'_1 x'_2 \ldots x'_n) = a_1 g^n + a_2 g^{n-1} + \ldots + a_n g \equiv 0 \pmod{2^{128}}.$$

Let us call a representation $a_i = x_i - x'_i$ as a *splitting* of a_i. There can be several splittings for a given a_i. Each of them induces two trajectories of intermediate values of h: the trajectory starting with a message $x_1 x_2 \ldots x_n$ and the trajectory starting with a message $x'_1 x'_2 \ldots x'_n$.

Let h_i and h'_i be the low bytes of h for the first and second trajectories, respectively before the additions $h + x_i$ and $h + x'_i$. Let us call a splitting *suitable* if

$$h_i + x_i < 256, \quad h'_i + x'_i < 256, \quad i = 1, 2, \ldots, n.$$

Let us evaluate the probability of existing a suitable splitting for a_i. We will assume that h_i, h'_i are realizations of independent random variables with uniform distribution over $\{0, 1, \ldots, 255\}$.

Bytes x_i and x'_i can take any value from intervals $\{0, \ldots, 255 - h_i\}$ and $\{0, \ldots, 255 - h'_i\}$, respectively. At the same time, the difference $x_i - x'_i$ takes value from the interval $\{-255 + h'_i, \ldots, 255 - h_i\}$.

Then a_i is in the interval $\{-255 + h'_i, \ldots, 255 - h_i\}$ with the probability

$$\Pr(-255 + h'_i \le a_i \le 255 - h_i) = \begin{cases} \Pr(h_i \le 255 - a_i), & a_i \ge 0, \\ \Pr(h'_i \le 255 - |a_i|), & a_i < 0, \end{cases}$$

that is equal to $1 - |a_i|/256$.

Thus, the probability that a suitable splitting exists for the whole sequence $a_1 a_2 \ldots a_n$ is the following:

$$p = \prod_{i=1}^{n} \left(1 - \frac{|a_i|}{256}\right).$$

This probability can be rather high. For example, $p \approx 1/25$ for the following sequence for $n = 18$:

$$(-64, 5, 73, 35, -53, 19, -10, -78, -44, 48, 61, -1, -80, 26, -22, 72, -31, 0).$$

Or, $p \approx 1/13$ for the following sequence for $n = 19$:

$$(-37, 34, -74, -4, -17, 33, -18, 21, 54, 33, -1, 58, -71, -13, -10, 11, -88, -19, 0).$$

Moreover, the probability can be increased if we change a strategy of finding suitable splittings. We can allow to modify splittings $a_1, \ldots, a_{i-1}$ that have been already built if it is impossible to find a splitting for a_i.

Table 4. Collisions of FNV-1a.

Message 1	Message 2
f1dd5921afd29cbd33b357184e8c	928ea41b7373792aae2bfa72ca64
eb18151b160aa95e0511357e158b58	ab775b3a7c7c7c7c7c3a5dc94e
f828e4070672220b195e0ddd2114a4c008	3638fa655d1b61e21419134803222bbb35
3a7a3a7a3a4a5a5a5a5a5a5a5a5a5a5a5a	51089c5e7fe7cc2d740b5f70b3cb5461824d
f4331cede51639057d05f80f1d6638b40b286f	eb2705051873321 16c611402081f1155326013
07160c2e0b700b1338ef6e63360419060507	10610bf23b0573e2317106176a171c6a4c6e
00ca0000cb00000000000000029092100d814	2d000158001b773a6364fc0905000000e90000

After finding a suitable splitting, we determine the sequences $(h_i), (h'_i)$. Then we determine the bytes $\tilde{x}_i, \tilde{x}'_i$ such that

$$h_i \oplus \tilde{x}_i = h_i + x_i, \quad h'_i \oplus \tilde{x}'_i = h'_i + x'_i, \quad i = 1, 2, ..., n.$$

It is important that there are no carries in high bytes in additions $h_i + x_i$, $h'_i + x'_i$; and $\tilde{x}_i, \tilde{x}'_i$ can be always found. Then a collision for FNV-1a is a pair of messages $\tilde{x}_1\tilde{x}_2...\tilde{x}_n$ and $\tilde{x}'_1\tilde{x}'_2...\tilde{x}'_n$.

It remains to say that the sequence (a_i) can be found using LLL algorithm. The algorithm is applied to the lattice defined by the basic vectors

$$\begin{aligned}
\mathbf{b}_1 &= (1, 0, ..., 0, g^{n-1} \bmod 2^{128}),\\
\mathbf{b}_2 &= (0, 1, ..., 0, g^{n-2} \bmod 2^{128}),\\
&\cdots\\
\mathbf{b}_n &= (0, 0, ..., 1, g^0 \bmod 2^{128}),\\
\mathbf{b}_{n+1} &= (0, 0, ..., 0, t2^{128}),
\end{aligned}$$

where t is a small integer. LLL finds a short basis of the lattice, i.e., vectors

$$v = \sum_{i=1}^{n+1} a_i \mathbf{b}_i$$

with small coordinate values. Let the last coordinate v equal to 0. Then

$$\sum_{i=1}^{n+1} a_i g^{n-i} \equiv 0 \pmod{2^{128}},$$

i.e., $(a_1, ..., a_n)$ is a required solution.

This problem was completely solved by fourteen teams (the most of them used a reduction to the problem FNV2). Some examples of collisions proposed by participants (in HEX format) are given in Table 4.

Problem "TwinPeaks2"

Formulation

Bob realized that his cipher from last year, TwinPeaks (NSUCRYPTO'2017) (Gorodilova et al. 2019), is not secure enough and modified it. He considerably

increased the number of rounds and made rounds more complicated. Bob's new cipher works as follows.

A message X is represented as a binary word of length 128. It is divided into four 32-bit words a, b, c, d and then the following round transformation is applied 48 times:

$$(a, b, c, d) \leftarrow (b, c, d, a \oplus S_3(S_1(b) \oplus S_2(b \wedge \neg c \oplus c \vee d) \oplus S_1(d))),$$

Here S_1, S_2, S_3 are secret permutations over 32-bit words; $\neg$, $\wedge$, $\vee$, $\oplus$ are binary bitwise "NOT", "OR", "AND", "XOR", respectively (the operations are listed in descending order of priority). The concatenation of the final a, b, c, d is the resulting ciphertext Y for the message X.

Agent Cooper again wants to read Bob's messages! He intercepted the ciphertext

$$Y = \texttt{DEB239852F1B47B005FB390120314478}$$

and also captured Bob's smartphone with the `TwinPeaks2` implementation! Here it is (https://nsucrypto.nsu.ru/olymp/2018/round/2/task/4). Now Cooper (and you too) can encrypt any messages with `TwinPeaks2` but still can not decrypt any. Help Cooper to decrypt Y.

Remarks. The ciphertext is given in hexadecimal notation, the first byte is DE.

Solution

Let F be the round transformation of `TwinPeaks2`:

$$F(a, b, c, d) = (b, c, d, a \oplus f(b, c, d)).$$

The encryption transformation is the composition of 48 copies of F, i.e., it can be written as F^{48}. Consequently, F^{-48} is the decryption transformation. Let

$$\tau(a, b, c, d) = (d, c, b, a).$$

Let us note that $f(b, c, d) = f(d, c, b)$. Then the composition of F, τ and F gives us τ:

$$F \circ \tau \circ F(a, b, c, d) = F(a \oplus f(b, c, d), d, c, b) = F(a \oplus f(d, c, b), d, c, b) = (d, c, b, a).$$

Hence,

$$F^{48} \tau F^{48} = \tau$$

or

$$F^{-48} = \tau F^{48} \tau^{-1} = \tau F^{48} \tau.$$

Thus, in order to decrypt Y one should write its 32-bit blocks in reverse order, encrypt the result and then reverse the order of the blocks again. The result will be a hexadecimal word, which gives us the desired message

`attacksgetbetter`.

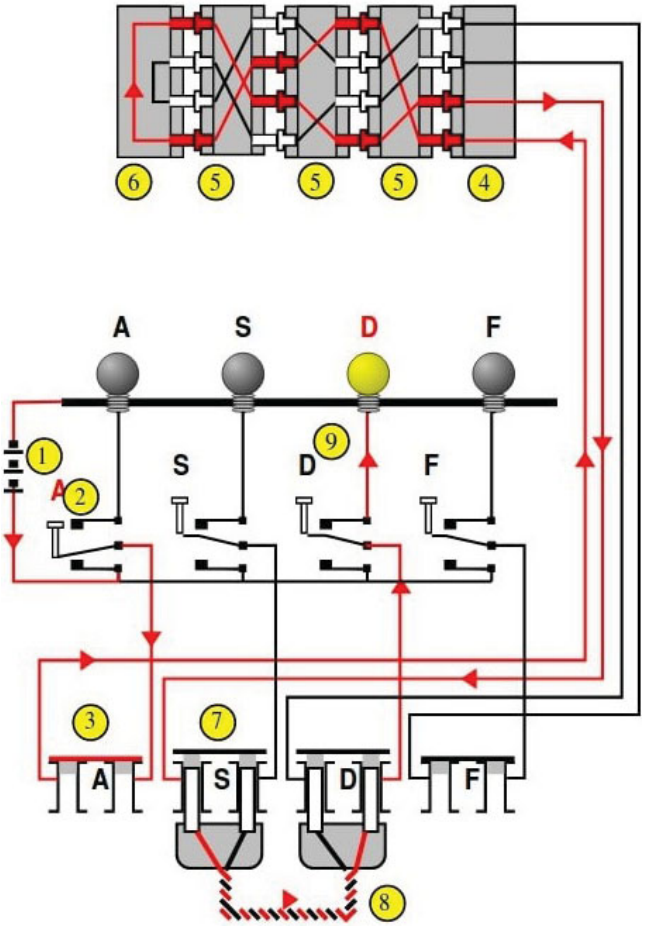


Figure 4. A simplified Enigma.

The best solution to the problem has been submitted by Carl Löndahl (Sweden), which not only provides a clean theoretical solution, but also proposes a slide attack on the cipher.

Problem "An Enigmatic challenge"

Formulation

The Enigma machine is a symmetric cipher famous for being used during the Second World War by the German military. Its internal structure comprises a 26-letter Latin alphabetic permutation, implemented as rotors. The machine used for this problem consists of three **rotors** and a **reflector**.

Figure 4 shows how a simplified Enigma machine works. The key components are the set of input switches (2)—which are reduced to 4 in the example but could have been 26 for the Latin alphabet—an input plugboard (3, 7, 8), three rotors (5), the reflector (6), and the output board (9).

The components have the following functionality:

- **Rotors:** a rotor (5) is a wheel with the upper-case alphabet in order on the rim and a hole for an axle. On both sides of a rotor are 26 electrical contacts each under a letter. Each contact on one side is wired to a contact on the other side at a different position. The rotor implements a one-to-one and onto function between the upper-case letters, where each letter is mapped to a different one (an irreflexive permutation).
- **Reflector:** the reflector (6) is positioned after the rotors and has contacts for each letter of the alphabet on one side only. The letters are wired up in pairs, so that an input current on a letter is reflected back to a different letter.

The input message: is permuted by the rotors, passes through the reflector, and then goes back through the rotors in the reverse order (as depicted in the figure). Finally, the light bulb indicates the encrypted letter. The plugboard plays no role in permuting the letter for this challenge, although it could have.

To prevent simple frequency analysis attack the right rotor rotates with every new input. After the right rotor completed a full rotation (after 26 letters were encrypted), the middle rotor rotates once. Similarly, after the middle rotor completes a full rotation (and the right rotor complete 676 rotations), the left rotor rotates once.[1]

Challenge: you will play the role of an attacker that knows the source of the plaintext to be encrypted. You are given a ciphertext corresponding to a plaintext taken from this known source which happens to be "Moby Dick" by Herman Melville, and you are asked to recover the plaintext. The plaintext consists only of trimmed capital letters with no punctuation marks and spaces and is contiguous. All letters are from the Latin alphabet. Extra information on the settings of the rotors is provided: the configuration of the first rotor is very close to the one used in the 1930 commercial version (that was `EKMFLGDQVZNTOWYHXUSPAIBRCJ`).

[1] This means that an input letter is processed, in order, by three permutation—right, middle, and left—reflected by the reflector, and processed once again, in order, by the inverse permutations corresponding to left, middle and right rotors before being output. Once the letter passes through a rotor, it is permuted with one position, the rotor's permutation is applied, and the result goes directly into the following rotor, which acts similarly.

Ciphertext:

RHSM	ZHXX	AOWW	ZTWQ	QQMB	CRZA	BARN	MLAV	MLSX	SPBA	ZTHG
YLGE	VGZG	KULJ	FLOZ	RQAW	YGAA	DCJB	YWBW	IYQQ	FAAO	RAGK
BGSW	OARG	EYSP	IKYE	LLUO	YCNH	HDBV	AFKD	HETA	ONNR	HXHE
BBRT	ROZD	XJCC	OMXR	PNSW	UAZB	TNJY	BANH	FGCS	GJWY	YTBV
VGLX	KUZW	PARO	NMXP	LDLZ	ICBK	XVSJ	NXCF	SOTA	AQYS	YZFX
MZDH	MSZI	ABAH	RFXT	FTPU	VWMC	PEXQ	NZVA	LMFX	BHKG	QGYS
BIYE	MEUE	PJNR	AVTL	JSUZ	PLHQ	MOUI	IQFD	HVXI	NOOJ	YJAF
WAVU	PVQA	FMKP	AHLK	XJYD	GITB	QSPK	CUZU	XPRK	MUJJ	YRJ

Link to "MobyDick" text file can be found in https://gist.githubusercontent.com/StevenClontz/4445774/raw/1722a289b665d940495645a5eaaad4da8e3ad4c7/mobydick.txt.

Solution

It is easy to observe that the left and middle rotors will not change for each block of 26 characters of the plaintext. From this point of view, we can regard the composition of permutations induced (in order) by the middle and left rotors, the reflector and as well as the inverses of the left and middle rotors, as one, fixed permutation. After the next 26 letters are processed, the middle rotor turns, and a distinct permutation is to be used for the incoming block of 26 letters. Due to the fact the challenge ciphertext is less than 676 characters, we do not bother with turning the left rotor.

To fix some notations, let π_i, L_i, M_i denote permutations defined on the set $\{\mathtt{A}, \ldots, \mathtt{Z}\}$. If $L : \{\mathtt{A}, \ldots, \mathtt{Z}\} \to \{\mathtt{A}, \ldots, \mathtt{Z}\}$ denotes the permutation defining the left rotor, by $L_i : \{\mathtt{A}, \ldots, \mathtt{Z}\} \to \{\mathtt{A}, \ldots, \mathtt{Z}\}$ and $L_i = L \circ Rot_i$, we represent the action of applying the left rotor over the alphabet, where Rot_i represents the alphabet's rotation by i. We use a similar notation for M_i, with i denoting each block of 26 letters to be processed. That is $i \in \{1, \ldots, \lceil \frac{|C|}{26} \rceil\}$, where $|C|$ denotes the length of the ciphertext (its number of characters). We also write

$$\pi_i = M_i^{-1} \circ L_0^{-1} \circ \rho \circ L_0 \circ M_i, \quad i \in \left\{1, \ldots, \left\lceil \frac{|C|}{26} \right\rceil \right\}.$$

The next step is to split the challenge ciphertext into blocks of 26 characters, and use the fact that for each block i, π_i acts as an oracle that returns the same value for the same input. We will correlate this with the information that is *a priori* given on the first rotor. Although we do not have its exact configuration, we use the fact that the unknown rotor is close to a known one (EKMFLGDQVZNTOWYHXUSPAIBRCJ—commercial Enigma 1930). The configuration used for this problem permutes four elements amongst the ones of the 1930 configuration and then applies a circular permutation of length four.

The permutation corresponding to the given right rotor of the commercial Enigma (1930) is the following:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
E	K	M	F	L	G	D	Q	V	Z	N	T	O	W	Y	H	X	U	S	P	A	I	B	R	C	J

Then we take the first block of 26 letters and obtain their inverses, minding the fact that the rotor shifts with one place to the left after we read one letter. Hence, for the first block we obtain:

RHSM	ZHXX	AOWW	ZTWQ	QQMB	CRZA	BA
UVAH	FOFL	VRDQ	TDNG	DQLA	BOIR	JJ

Now we remark on a "distance-preserving" property: if the distance between identical characters returned by π_i (the input to the Right rotors) is ℓ, then it maintained in the original plaintext. As an example, the group `ZHXX` in the first block of ciphertext has been obtained for the group `FOFL` and we note a distance of 2 (`F` $\rightarrow$ `O` $\rightarrow$ `F`) between `F` and `F`. This means that an alphabetical distance of 2 exists between the corresponding letters of the plaintext. More precisely, if:

$$\pi_1(R(x)) = \pi_1(R'(y)),$$

where R' is obtained by shifting R with ℓ elements, then the character y is at a distance of ℓ from the character x (but in the opposite sense). Based on this observation, the solution is to identify such pairs inside a block and record the distance between them. As four elements are permuted in the real configuration of the rotor, false positives will appear.

After the colliding characters per block, say in position i and j, have been identified and their distance recorded, say the distance is ℓ, one will simply write a script that will pass through the given plaintext (after removing the non-alphabetic characters) and identify the sequence (matching the length of the ciphertext) where the distance between the characters in position i and j is ℓ.

Finally, the plaintext that is to be recovered is:

> ALREADY we are boldly launched upon the deep; but soon we shall be lost in its unshored, harbourless immensities. Ere that come to pass; ere the Pequod'a weedy hull rolls side by side with the barnacled hulls of the leviathan; at the outset it is but well to attend to a matter almost indispensable to a thorough appreciative understanding of the more special leviathanic revelations and allusions of all sorts which are to follow.

Finally, eight teams completely solved the problem. Note, that many teams used a simple method that almost completely determined the plaintext. It is based on the fact that no letter from the plaintext gets mapped to

the same letter in the ciphertext using Enigma. But this approach gives two possible solutions and does not allow one to prove that one of them is not correct.

Problem "Orthogonal arrays" (unsolved)

Formulation

Orthogonal arrays are closely connected with cryptographic Boolean functions. Namely, supports of correlation immune functions give orthogonal arrays when their elements are written as the rows of an array.

Given three positive integers n, t, and λ such that $t < n$, we call a $\lambda 2^t \times n$ binary array (i.e., matrix over the two-element field) a $t-(2, n, \lambda)$ *orthogonal array* if in every subset of t columns of the array, every (binary) t-tuple appears in exactly λ rows. t is called the strength of this orthogonal array.

Find a $4-(2, 11, \lambda)$ orthogonal array with minimal value of λ.

Solution

The best known answer to this question is $\lambda = 8$ (Picek et al. 2015), but it is unknown whether there exists a $4-(2, 11, \lambda)$ orthogonal array for $\lambda < 8$. This open problem remains unsolved. Participants suggested several ideas.

The most interesting one was proposed by Aleksei Udovenko (University of Luxembourg). His study starts with the Nordstrom—Robinson code (that is, the Kerdock code of length 16 and size 256, whose dual distance is the minimum distance of the Preparata code, that is 6, which gives a strength of the orthogonal array (OA) equal to 5). Only the codewords with the first element equal to zero are kept and their coordinate at 0 is deleted, which makes size 128, length 15 and strength 4. Then three columns are erased from the OA, which does not reduce the strength, and the resulting OA provides a solution to the problem with $\lambda = 8$. It is then shown (by using known results) that, for any solution to the problem, λ is at least 6. This is interesting. The solution found is written in the form $(x, F(x))$ where F is a quadratic (7, 4)-function. Its determination allows one to determine the 4-th order correlation immune function whose support is this OA. This is an 11-variable Boolean function of algebraic degree 5. Then the annihilators of this function are studied. It is shown that the function has a linear annihilator (and has then algebraic immunity 1). After an observation on the impossibility of extending a solution which would have $\lambda \leq 7$, the Xiao—Massey characterization of OA is proved again in different terms. It is also shown that any affine annihilator of a t-th order correlation immune function must be t-resilient which is a nice observation. A computer search is made with Integer Linear Programing

showing that any 4-th order correlation immune function having an affine annihilator should have weight at least 128, which is a nice observation. This nice work concludes with open questions.

Another good solution was given by the team of Evgeniya Ishchukova, Vyacheslav Salmanov, and Oksana Shamilyan (Southern Federal University, Russia). They first studied the maximum value of n, given t, for small values of t. Then an algorithm was designed which reduces the search to solutions having some symmetries observed in smaller values of t and n. Finally, a solution was given with $\lambda = 8$ which is the coset of a linear code of length $n = 11$ and dimension $k = 7$ (and therefore 128 codewords), with dual distance 5, and the corresponding function is then indeed 4th order correlation immune giving a $4-(2, 11, \lambda)$ orthogonal array. Unfortunately, the question whether 128 is minimal was not addressed.

Problem "Sylvester matrices" (unsolved)

Formulation

Sylvester matrices play a role in security since they are connected with topics like secret sharing and MDS codes constructed with cellular automata.

Consider two univariate polynomials over the two-element field, $P_1(x)$ of degree m and $P_2(x)$ of degree n, where $P_1(x) = a_m x^m + \cdots + a_0$ and $P_2(x) = b_n x^n + \cdots + b_0$. The *Sylvester matrix* is an $(m+n) \times (m+n)$ matrix formed by filling the matrix beginning with the upper left corner with the coefficients of $P_1(x)$, then shifting down one row and one column to the right and filling in the coefficients starting there until they hit the right side. The process is then repeated for the coefficients of $P_2(x)$. All the other positions are filled with zero.

Let $n > 0$, $m > 0$. Prove whether there exist $(m+n) \times (m+n)$ invertible Sylvester matrices whose inverses are Sylvester matrices as well.

Example. For $m = 4$ and $n = 3$, the Sylvester matrix is the following:

$$\begin{pmatrix} a_4 & a_3 & a_2 & a_1 & a_0 & 0 & 0 \\ 0 & a_4 & a_3 & a_2 & a_1 & a_0 & 0 \\ 0 & 0 & a_4 & a_3 & a_2 & a_1 & a_0 \\ b_3 & b_2 & b_1 & b_0 & 0 & 0 & 0 \\ 0 & b_3 & b_2 & b_1 & b_0 & 0 & 0 \\ 0 & 0 & b_3 & b_2 & b_1 & b_0 & 0 \\ 0 & 0 & 0 & b_3 & b_2 & b_1 & b_0 \end{pmatrix}$$

Solution

We are pleased to say that three teams completely solved this problem! They are Alexey Chilikov (Bauman Moscow State Technical University, Russia), the team of Radu Caragea, Madalina Bolboceanu and Miruna Rosca (Bitdefender, Romania), and the team of Samuel Tang and Harry Lee (Hong Kong). Here we present the main idea for the solution.

Case 1: $m \le n$. Let $P_1(x) = x^m$ and $P_2(x) = x^n + 1$. Then their Sylvester matrix is the following:

$$\left(\begin{array}{c|c|c} \mathbf{I}_n & & 0_{n\times m} \\ \hline \mathbf{I}_m & 0_{m\times(n-m)} & \mathbf{I}_m \end{array}\right),$$

where $\mathbf{I}_k$ denotes the $k \times k$ identity matrix; and $0_{k\times\ell}$ is the $k \times \ell$ zero matrix. Taking all operations over the two-element field, it is clear that

$$\left(\begin{array}{c|c|c} \mathbf{I}_n & & 0_{n\times m} \\ \hline \mathbf{I}_m & 0_{m\times(n-m)} & \mathbf{I}_m \end{array}\right) \cdot \left(\begin{array}{c|c|c} \mathbf{I}_n & & 0_{n\times m} \\ \hline \mathbf{I}_m & 0_{m\times(n-m)} & \mathbf{I}_m \end{array}\right) = \left(\begin{array}{c|c} \mathbf{I}_n & 0_{n\times m} \\ \hline 0_{m\times n} & \mathbf{I}_m \end{array}\right) = \mathbf{I}_{m+n}.$$

Thus, the considered Sylvester matrix is an involutory matrix. Therefore, its inverse is the Sylvester matrix as well.

Case 2: $m \ge n$. Assume that the inverse of the Sylvester matrix of $P_1(x)$ and $P_2(x)$ is also the Sylvester matrix for two polynomials over the two-element field, say $Q_1(x) = c_p x^p + c_{p-1}x^{p-1} + \cdots + c_0, Q_2(x) = d_q x^q + d_{q-1}x^{q-1} + \cdots + d_0$, of degrees $p > 0$ and $q > 0$, respectively, which satisfy $p + q = m + n$. The product of Sylvester matrices which correspond to $P_1(x), P_2(x)$ and $Q_1(x), Q_2(x)$ is equal to $\mathbf{I}_{m+n}$, in particular

$$\left(\begin{array}{ccccccc} a_m & a_{m-1} & \dots & a_0 & & & \\ & a_m & a_{m-1} & \dots & a_0 & & \\ & & \ddots & \ddots & & \ddots & \\ & & & a_m & a_{m-1} & \dots & a_0 \\ \hline b_n & b_{n-1} & \dots & b_0 & 0 & \dots & 0 \\ \hline & b_n & b_{n-1} & \dots & b_0 & & \\ & & \ddots & \ddots & & \ddots & \\ & & & b_n & b_{n-1} & \dots & b_0 \end{array}\right) \cdot \begin{pmatrix} c_p \\ 0 \\ \vdots \\ 0 \\ d_q \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{F}_2^{m+n}.$$

The condition $q > n$ implies $b_n c_p = 0$, but $b_n = c_p = 1$ since the polynomials $P_2(x)$ and $Q_1(x)$ have degrees n and p, respectively. Therefore, it must hold $q \le n$. Since $Q_2(x)$ has degree q, then $d_q = 1$ and $(1 + b_{n-q}) = b_{n-q+1} = b_{n-q+2} = \ldots = b_{n-q+\min\{q,m-1\}} = 0$. From $b_n = 1$ it follows that $\min\{q, m-1\} < q$, that is $m \le q$. Finally, we get $m \le q \le n < m$ that is a contradiction.

Thus, in the case $m \leq n$ there exist invertible Sylvester matrices whose inverse are Sylvester matrices as well but for $m > n$ it does not hold.

Problem "Disjunct matrices" (unsolved)

Formulation

Disjunct matrices are used in some key distribution protocols for traitor tracing. Disjunct matrices (DM) are a particular kind of binary matrices which have been applied to solve the non-adaptive group testing (NAGT) problem, where the task is to detect any configuration of t defectives out of a population of N items. Traditionally, the methods used to construct DM leverage on error-correcting codes and other related algebraic techniques.

Let $A = (x_1^\top, x_2^\top, ..., x_N^\top)$ be an $M \times N$ binary matrix. Then, A is called *t-disjunct* if, for all subsets of t columns $S = \{x_{i_1}, ..., x_{i_t}\}$, and for all remaining columns $x_j \notin S$, it holds that

$$\mathrm{supp}(x_j) \not\subseteq \bigcup_{k=1}^{t} \mathrm{supp}(x_{i_k}),$$

where $\mathrm{supp}(x)$ denotes the set of coordinate positions of a binary vector x with 1s.

In other words, a matrix A is t-disjunct if for every subset S of t columns the support of any other column is not contained in the union of the supports of the columns in S.

Prove what is the minimum number of rows in a 5-disjunct matrix.

Solution

We must admit that the formulation of the problem did not include the condition which makes this problem non-trivial. The condition is that the number of columns must be greater than the number of rows. This formulation comprises practical significance and has the following equivalent form: given t, when does there exist a t-disjunct algorithm better than the trivial one that tests each item individually? Readers may find details regarding non-adaptive group testing (NAGT) problem together with known results and mentioned formulations in Shangguan and Ge (2016).

The solution of the originally stated problem is 6—consider the 6×6 identity matrix. This solution was discovered by several participants. However some participants (Alexey Chilikov from Bauman Moscow State Technical University, Aleksei Udovenko from University of Luxembourg, and the team of Henning Seidler and Katja Stumpp from Technical

University of Berlin) obtained bounds for the number of rows depending the parameter t and the number of columns.

Winners of the Olympiad

Figure 5 shows NSUCRYPTO winners from 2014 to 2018. Here we list information about the winners of NSUCRYPTO'2018 in Tables 5–10.

Figure 5. Winners of NSUCRYPTO from 2014 to 2018.

Table 5. Winners of the first round in school section A ("school student").

Place	Name	Country, City	School	Scores
1	Borislav Kirilov	Bulgaria, Sofia	The First Private Mathematical Gymnasium	22
1	Alexey Lvov	Russia, Novosibirsk	Gymnasium 6	21
1	Razvan Andrei Draghici	Romania, Craiova	National College Fratii Buzesti	20
2	Gorazd Dimitrov	Macedonia, Skopje	Yahya Kemal College	18
3	Ivan Baksheev	Russia, Novosibirsk	Gymnasium 6	16
3	Artem Ismagilov	Russia, Yekaterinburg	The Specialized Educational and Scientific Center UrFU	16
3	Bogdan Circeanu	Romania, Craiova	National College Fratii Buzesti	16
3	Ruxandra Icleanu	Romania, Craiova	Tudor Vianu National College of Computer Science	15
Diploma	Sofya Gorbunova	Russia, Yekaterinburg	The Specialized Educational and Scientific Center UrFU	13
Diploma	Kirill Poltoradnev	Russia, Yekaterinburg	The Specialized Educational and Scientific Center UrFU	13
Diploma	Tudor Moga	Romania, Brasov	Grigore Moisil National College of Computer Science	11
Diploma	Markas Cerniauskas	Lithuania, Kaunas	Kaunas Technology University Gymnasium	11
Diploma	Mircea-Costin Preoteasa	Romania, Bucharest	Tudor Vianu National College of Computer Science	11
Diploma	Kirill Tugolukov	Russia, Ulan-Ude	School 19, Olympiad training center ENTER	11

Table 6. Winners of the first round, section B (in the category "university student").

Place	Name	Country, City	University	Scores
1	Maxim Plushkin	Russia, Moscow	Lomonosov Moscow State University	25
2	Robert Koprinkov	Netherlands, Nijmegen	Radboud University	20
2	Irina Slonkina	Russia, Moscow	National Research Nuclear University MEPhI	18
2	Marc Houben	Belgium, Leuven	KU Leuven	18
3	Dheeraj M Pai	India, Chennai	Indian Institute of Technology Madras	17
3	Alexander Grebennikov	Russia, Saint Petersburg	Saint Petersburg State University	16
3	Roman Lebedev	Russia, Novosibirsk	Novosibirsk State University	16
3	Dianthe Bose	India, Chennai	Chennai Mathematical Institute	15
3	Ivan Sutormin	Russia, Novosibirsk	Novosibirsk State University	15
Diploma	Harikumar Krishnamurthy	India, Chennai	Indian Institute of Technology Madras	14
Diploma	Roman Tarasov	Russia, Odintsovo	Higher School of Economics	13
Diploma	Sander Suverkropp	Netherlands, Wageningen	Radboud University	13
Diploma	Saeed Odak	Iran, Tehran	Khajeh Nasir Toosi University of Technology	13
Diploma	Thijs van Loenhout	Netherlands, Nijmegen	Radboud University	12
Diploma	Daniil Gurev	Russia, Novosibirsk	Novosibirsk State University	12
Diploma	Neha Rino	India, Chennai	Chennai Mathematical Institute	12
Diploma	Kristina Volyakova	Russia, Yaroslavl	Yaroslavl State University	12

Table 7. Winners of the first round, section B (in the category "professional").

Place	Name	Country, City	Organization	Scores
1	Alexey Udovenko	Luxembourg, Luxembourg	University of Luxembourg	23
3	Henning Seidler	Germany, Berlin	TU Berlin	16
Diploma	Alexey Chilikov	Russia, Moscow	Bauman Moscow State Technical University	14
Diploma	Samuel Tang	Hong Kong, Hong Kong	Blocksquare Limited	12
Diploma	Amedeo Sgueglia	United Kingdom, London	London School of Economics and Political Science	12
Diploma	Samad Alaamati	Iran, Tehran	American Society for Industrial Security	12

Table 8. Winners of the second round (in the category "school student").

Place	Names	Country, City	School	Scores
Diploma	Brian Ncube	Zimbabwe, Hwange	Hwange High School	7

Table 9. Winners of the second round (in the category "university student").

Place	Name	Country, City	University	Scores
1	Maxim Plushkin	Russia, Moscow	Lomonosov Moscow State University	43
2	Irina Slonkina	Russia, Moscow	National Research Nuclear University MEPhI	38
2	Dmitry Lavrenov, Egor Lavrenov, Uladzimir Paprotski	Belarus, Minsk	Belarusian State University	37
3	Thanh Nguyen Van, Tuong Nguyen Van, Dinh Ton	Vietnam, Ho Chi Minh City	Ho Chi Minh City University of Technology, University of Science	37
3	Ngoc Ky Nguyen, Phuoc Nguyen Ho Minh, Danh Nam Tran	Vietnam, Ho Chi Minh City; France, Paris	Ho Chi Minh City Pedagogical University, Ho Chi Minh City University of Technology, Ecole Normale Superieure	34

(*continued*)

Table 9. Continued.

Place	Name	Country, City	University	Scores
3	Dianthe Bose, Neha Rino	India, Chennai	Chennai Mathematical Institute	33
3	Roman Lebedev, Vladimir Sitnov, Alexander Tkachev	Russia, Novosibirsk	Novosibirsk State University	30
3	Mikhail Sorokin, Darya Frolova, Vladimir Bobrov	Russia, Moscow	National Research Nuclear University MEPhI	29
3	Andrey Kalachev, Danil Cherepanov, Alexey Radaev	Russia, Moscow	Bauman Moscow State Technical University	27
Diploma	Saveliy Skresanov	Russia, Novosibirsk	Novosibirsk State University	21
Diploma	Harikumar Krishnamurthy, Aditya Pradeep, Dheeraj M Pai	India, Chennai	Indian Institute of Technology Madras	21

Table 10. Winners of the second round (in the category "professional").

Place	Names	Country, City	Organization	Scores
1	Alexey Udovenko	Luxembourg, Luxembourg	University of Luxembourg	72
2	Evgeniya Ishchukova, Vyacheslav Salmanov, Oksana Shamilyan	Russia, Taganrog	Southern Federal University	46
3	Henning Seidler, Katja Stumpp	Germany, Berlin	Berlin Technical University	40
3	Carl Londahl	Sweden, Karlskrona	TrueSec AB	38
3	Alexey Chilikov	Russia, Moscow	Bauman Moscow State Technical University	38
Diploma	Duc Tri Nguyen, Quan Doan, Quoc Bao Nguyen	Vietnam, Ho Chi Minh city	CERG at George Mason University, E-CQURITY, Ho Chi Minh City University of Technology	37
Diploma	Radu Caragea, Miruna Rosca, Madalina Bolboceanu	Romania, Bucharest	Bitdefender	32
Diploma	Mikhail Polyakov, Mikhail Tsvetkov, Victoria Vlasova	Russia, Moscow	Bauman Moscow State Technical University	30
Diploma	Harry Lee, Samuel Tang	Hong Kong, Hong Kong	Blocksquare Limited, Hong Kong University of Science and Technology	29
Diploma	Lars Haulin	Sweden, Uppsala	–	28
Diploma	Sergey Titov, Kristina Geut	Russia, Yekaterinburg	Ural State University of Railway Transport	22
Diploma	Khai Hanh Tang, Neng Zeng, Thu Hien Chu Thi	Singapore, Singapore	Ho Chi Minh City Pedagogical University, Nanyang Technological University	21

About the authors

Anastasiya Gorodilova is a researcher at the Sobolev Institute of Mathematics; an Assistant Professor at Novosibirsk State University; a researcher at the Mathematical center in Akademgorodok. She is interested in cryptographic Boolean functions, APN functions, bent functions, symmetric cryptography, combinatorics, and algebra.

Sergey Agievich is the head of the IT Security Research Laboratory of the Research Institute for Applied Problems of Mathematics and Informatics (Belarusian State University). His research interests include Boolean functions in cryptography, cryptographic algorithms and

protocols, enumerative and asymptotic combinatorics, exponential sums and systems of polynomial equations.

Claude Carlet is a Professor emeritus of mathematics at the University of Paris 8, Laboratory LAGA; a member of the Department of Informatics of the University of Bergen. His research interests include algebra, coding theory and cryptography, cryptographic Boolean functions.

Xiang-dong Hou is a Professor at the faculty of Mathematics and Statistics in the University of South Florida. His research areas are algebra, number theory, coding theory and cryptography, combinatorics, and topology.

Valeria Idrisova is a researcher at the Sobolev Institute of Mathematics and Novosibirsk State University; a researcher at the Mathematical center in Akademgorodok. Her research interests includes vectorial Boolean functions, APN permutations, block ciphers, and side-channel attacks.

Nikolay Kolomeec is a researcher at the Sobolev Institute of Mathematics; an Assistant at Novosibirsk State University; a researcher in Mathematical center in Akademgorodok. Research areas are pseudorandom sequences, cryptographic Boolean functions, bent functions, and discrete mathematics.

Alexandr Kutsenko is a PhD student at the Department of Mathematics and Mechanics in Novosibirsk State University; a researcher at the Mathematical center in Akademgorodok. His research interests include mathematical problems of quantum cryptography and cryptographic Boolean functions.

Luca Mariot is a postdoctoral researcher at the University of Milano-Bicocca, and a member of the BiS Lab (Bicocca Security Lab). His research interests include natural computing models and techniques for cryptography.

Alexey Oblaukhov is a PhD student at the Sobolev Institute of Mathematics; an assistant at Novosibirsk State University; a researcher at the Mathematical center in Akademgorodok. His research interests are blockchain technologies, cryptography, and discrete mathematics.

Stjepan Picek is an assistant professor in the Cyber Security research group of the faculty of Electrical Engineering, Mathematics and Computer Science at Delft University of Technology. His main research interests are at the intersection of cryptography, cybersecurity, evolutionary computation, and machine learning.

Bart Preneel is a full professor in the research group COSIC of the Electrical Engineering Department of KU Leuven; a director of the International Association for cryptologic Research. Main interests are cryptography and information security. His research focuses on cryptographic algorithms and protocols as well as their applications to computer and network security and mobile communications.

Razvan Rosie is a researcher at University of Luxembourg. His research interests are in data security, focusing on constructions and applications of primitives in the area of public-key cryptography.

Natalia Tokareva is a senior researcher in the Sobolev Institute of Mathematics; associate professor at Novosibirsk State University; a head of the Laboratory of Cryptography JetBrains Research; a researcher at the Mathematical center in Akademgorodok. Her research interests include Boolean functions in cryptography, bent functions, block and stream ciphers, cryptanalysis, coding theory, combinatorics, and algebra.

Funding

The paper was supported by the Russian Ministry of Science and Education (the 5–100 Excellence Programme and the Project no. 1.13559.2019/13.1), by the Russian Foundation for Basic Research (project nos. 18-07-01394, 18-31-00479, and 18-31-00374), by the program of fundamental scientific researches of the SB RAS no. I.5.1, project no. 0314-2019-0017, by JetBrains Research, Novosibirsk, Russia.

ORCID

Anastasiya Gorodilova http://orcid.org/0000-0002-4986-434X
Sergey Agievich http://orcid.org/0000-0002-9413-8574
Claude Carlet http://orcid.org/0000-0002-6118-7927
Valeria Idrisova http://orcid.org/0000-0002-3013-5446
Nikolay Kolomeec http://orcid.org/0000-0003-4367-3507
Alexandr Kutsenko http://orcid.org/0000-0001-5006-1248
Luca Mariot http://orcid.org/0000-0003-3089-6517
Alexey Oblaukhov http://orcid.org/0000-0002-3613-4647
Stjepan Picek http://orcid.org/0000-0001-7509-4337
Bart Preneel http://orcid.org/0000-0003-2005-9651
Natalia Tokareva http://orcid.org/0000-0002-4343-7048

References

Agievich, S., A. Gorodilova, V. Idrisova, N. Kolomeec, G. Shushuev, and N. Tokareva. 2017. Mathematical problems of the second international student's Olympiad in cryptography. *Cryptologia* 41 (6):534–65. doi:10.1080/01611194.2016.1260666.

Agievich, S., A. Gorodilova, N. Kolomeec, S. Nikova, B. Preneel, V. Rijmen, G. Shushuev, N. Tokareva, and V. Vitkup. 2015. Problems, solutions and experience of the first international student's Olympiad in cryptography. *Prikladnaya Diskretnaya Matematika (Applied Discrete Mathematics)* 29 (3):41–62. doi:10.17223/20710410/29/4

Agievich, S., V. Marchuk, A. Maslau, and V. Semenov. 2016. Bash-f: Another LRX sponge function. Report 2016/587, Cryptology ePrint Archive.

Geut, K., K. Kirienko, P. Sadkov, R. Taskin, and S. Titov. 2017. On explicit constructions for solving the problem "A secret sharing". *Prikladnaya Diskretnaya Matematika. Prilozhenie* 10:68–70. doi:10.17223/2226308X/10/29.

Gorodilova, A., S. Agievich, C. Carlet, E. Gorkunov, V. Idrisova, N. Kolomeec, A. Kutsenko, S. Nikova, A. Oblaukhov, S. Picek, et al. 2019. Problems and solutions of the Fourth International Students Olympiad in Cryptography (NSUCRYPTO). *Cryptologia* 43 (2):138–74. doi:10.1080/01611194.2018.1517834.

Oblaukhov, A. 2019. A lower bound on the size of the largest metrically regular subset of the Boolean cube. *Cryptography and Communications* 11 (4):777–91. doi:10.1007/s12095-018-0326-1.

Picek, S., S. Guilley, C. Carlet, D. Jakobovic, and J. F. Miller. 2015. Evolutionary approach for finding correlation immune Boolean functions of order t with minimal Hamming weight. In *Theory and practice of natural computing. Lecture notes in computer science*, ed. A. H. Dediu, L. Magdalena, & C. Martn-Vide, vol. 9477, 71–82. Cham, Switzerland: Springer.

Tokareva, N., A. Gorodilova, S. Agievich, V. Idrisova, N. Kolomeec, A. Kutsenko, A. Oblaukhov, and G. Shushuev. 2018. Mathematical methods in solutions of the problems from the Third International Students' Olympiad in Cryptography. *Prikladnaya diskretnaya matematika* 40:34–58. doi:10.17223/20710410/40/4.

Shangguan, C., and G. Ge. 2016. New bounds on the number of tests for disjunct matrices. *IEEE Transactions on Information Theory* 62 (12):7518–21. doi:10.1109/TIT.2016.2614726.

Cryptography and Communications
https://doi.org/10.1007/s12095-020-00438-y

The group of automorphisms of the set of self-dual bent functions

Aleksandr Kutsenko[1,2]

Received: 25 August 2019 / Accepted: 6 May 2020 / Published online: 11 June 2020

Abstract
A bent function is a Boolean function in even number of variables which is on the maximal Hamming distance from the set of affine Boolean functions. It is called self-dual if it coincides with its dual. It is called anti-self-dual if it is equal to the negation of its dual. A mapping of the set of all Boolean functions in n variables to itself is said to be isometric if it preserves the Hamming distance. In this paper we study isometric mappings which preserve self-duality and anti-self-duality of a Boolean bent function. The complete characterization of these mappings is obtained for $n \geqslant 4$. Based on this result, the set of isometric mappings which preserve the Rayleigh quotient of the Sylvester Hadamard matrix, is characterized. The Rayleigh quotient measures the Hamming distance between bent function and its dual, so as a corollary, all isometric mappings which preserve bentness and the Hamming distance between bent function and its dual are described.

Keywords Boolean functions · Self-dual bent · Isometric mappings · The group of automorphisms · The Rayleigh quotient

1 Introduction

The term "bent function" was introduced by Oscar Rothaus in the 1960s [18]. It is known [21], that at the same time Boolean functions with maximal nonlinearity were also studied in the Soviet Union. The term *minimal function*, which is actually a counterpart of a bent function, was proposed by the Soviet scientists Eliseev and Stepchenkov in 1962.

This article belongs to the Topical Collection: *Boolean Functions and Their Applications IV*
Guest Editors: Lilya Budaghyan and Tor Helleseth

The author was supported by the Russian Foundation for Basic Research (projects no. 18-07-01394, 20-31-70043), the study was supported within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and Laboratory of Cryptography JetBrains Research.

✉ Aleksandr Kutsenko
Alexandrkutsenko@bk.ru

[1] Sobolev Institute of Mathematics, Novosibirsk, Russia

[2] Novosibirsk State University, Novosibirsk, Russia

Bent functions have connections with such combinatorial objects as Hadamard matrices and difference sets. Since bent functions have maximum Hamming distance to linear structures and affine functions they deserve attention for practical applications in symmetric cryptography, in particular, for block and stream ciphers. We refer to the survey [3] and monographies of Mesnager [17] and Tokareva [21] for more information concerning known results and open problems related to bent functions.

For each bent function, its dual bent function is uniquely defined. More information about properties of dual bent functions one can find in work [3]. A bent function that coincides with its dual is called self-dual. There are a number of papers devoted to open problems including characterization and description of the class of self-dual bent functions.

All equivalence classes of self-dual bent functions in 2, 4, and 6 variables and all quadratic self-dual bent functions in 8 variables with a respect to a restricted form of affine transformation which preserves self-duality were given in [2]. Further, equivalence classes of cubic self-dual bent functions in 8 variables with respect to the mentioned above restricted form of affine transformation one can find in [7]. In [8] a classification of quadratic self-dual bent functions was obtained. The upper bound for the cardinality of the set of self-dual bent functions was given in [9]. In [19] Sok et al. discovered a connection between quaternary self-dual bent functions and self-dual bent Boolean functions. New constructions of self-dual bent functions were presented in [13, 16]. The complete Hamming distance spectrum between self-dual Maiorana–McFarland bent functions was obtained in [11]. Iterative constructions and metrical properties, in particular, sets of Boolean functions which are maximally distant from the sets of self-dual and anti-self-dual bent functions and also the questions concerning metrical regularity of the sets of self-dual and anti-self-dual bent functions, were studied in [12].

Study of automorphism groups of mathematical objects deserve attention since these groups are closely connected with the structure of the objects. There exists an essential generally non-trivial question: how groups of automorphisms of two mathematical objects, one of which is embedded to another one, are related.

The group of automorphisms of the set of bent functions was completely characterized by Tokareva in [20]: it was proved that each isometric mapping of the set of Boolean functions in n variables to itself preserving the class of bent functions is a combination of an affine transformation of coordinates and a shift by an affine function. The said group is a semidirect product of the affine group $\mathrm{GA}(n, \mathbb{F}_2)$ and $\mathbb{F}_2^{n+1}$. A natural question arises how the automorphism group of the set of self-dual bent functions is connected with the group of automorphisms of the set of bent functions.

As it was mentioned, in papers [2, 7] the classification of self-dual bent functions based on the restricted form of affine equivalence preserving self-duality that forms the extended orthogonal group, was proposed. We study a question whether there exist other isometric mappings of Boolean functions in n variables to itself which preserve the class of self-dual bent function. In this paper, we prove that there are no other mappings satisfying such a property, thus obtaining a characterization of the group of automorphisms of the set of self-dual bent functions.

In this paper we study isometric mappings of the set of all Boolean functions in $n \geqslant 4$ variables to itself which preserve self-duality and anti-self-duality of a Boolean function. The complete characterization of these mappings is obtained. It is proved that every such mapping has form

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x \rangle \oplus d,$$

where L is a $n \times n$ orthogonal binary matrix, $c \in \mathbb{F}_2^n$, c has even Hamming weight, $d \in \mathbb{F}_2$. Based on this result, the set of isometric mappings which preserve the Rayleigh quotient of the Sylvester Hadamard matrix of every Boolean function is obtained. As a corollary all isometric mappings which preserve bentness and the Hamming distance between bent function and its dual are given.

The work has the following structure: basic definitions and notions concerning isometric mappings and groups of automorphisms are in the Sections 2 and 3. In Section 4 required material on sign functions of (anti-)self-dual bent function, which is directly used throughout the paper, is given. In Section 5 we characterize isometric mappings preserving self-duality (Theorem 1) and prove that isometric mapping preserves self-duality if and only if it preserves anti-self-duality (Proposition 2). In Section 6 isometric mappings which define bijections between the sets of self-dual and anti-self-dual bent functions (Theorem 2) are characterized. Section 7 is devoted to the Rayleigh quotient of a Boolean function and isometric mappings which preserve it (Theorem 3) and change its sign (Theorem 4) for every Boolean function. In Section 8 we summarize results from this paper (Theorems 6 and 7), the group of automorphisms of (anti-)self-dual bent functions is provided in Theorem 8. The conclusion is in Section 9.

2 Preliminaries

Let $\mathbb{F}_2^n$ be a set of binary vectors of length n.

A *Boolean function* f in n variables is any map from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. The set of Boolean functions in n variables is denoted by $\mathcal{F}_n$.

The (0, 1)-sequence defined by $(f(\mathbf{v}_0), f(\mathbf{v}_1), ..., f(\mathbf{v}_{2^n-1}))$ is called the *truth table* of $f \in \mathcal{F}_n$, where

$$\begin{aligned}
&\mathbf{v}_0 = (0, 0, ..., 0) \in \mathbb{F}_2^n \\
&\mathbf{v}_1 = (0, 0, ..., 0, 1) \in \mathbb{F}_2^n \\
&\vdots \\
&\mathbf{v}_{2^n-1} = (1, 1, ..., 1) \in \mathbb{F}_2^n,
\end{aligned}$$

ordered by lexicographical order.

The *sign function* F of a Boolean function $f \in \mathcal{F}_n$ is a real-valued function $F(x) = (-1)^{f(x)}$, $x \in \mathbb{F}_2^n$. Obviously, we have $(-1)^{f(x)} = 1 - 2f(x)$ for any $x \in \mathbb{F}_2^n$. We will denote the sign function by $F = (-1)^f$ and refer to it as to a vector $F = \left((-1)^{f(\mathbf{v}_0)}, (-1)^{f(\mathbf{v}_1)}, ..., (-1)^{f(\mathbf{v}_{2^n-1})}\right)$ from the set $\{\pm 1\}^{2^n}$ (it is also known as a $(1, -1)$-*sequence* of the function $f \in \mathcal{F}_n$, see [4]).

The sign $\oplus$ denotes a sum modulo 2. For $x, y \in \mathbb{F}_2^n$ denote $\langle x, y\rangle = \bigoplus_{i=1}^{n} x_i y_i$. Boolean function $f \in \mathcal{F}_n$ which can be represented in the form $f(x) = \langle a, x\rangle \oplus a_0, x \in \mathbb{F}_2^n$, where $a \in \mathbb{F}_2^n, a_0 \in \mathbb{F}_2$, is called *affine*. Two Boolean functions $f, g \in \mathcal{F}_n$ are said to be *affinely equivalent* if $g(x) = f(Ax \oplus b) \oplus \langle b, x\rangle \oplus d$, where $b, c \in \mathbb{F}_2^n, d \in \mathbb{F}_2$ and A is a $n \times n$ nonsingular binary matrix. If no such transformation exists, then f, g are called *inequivalent*.

The *Hamming weight* $\mathrm{wt}(x)$ of the vector $x \in \mathbb{F}_2^n$ is the number of nonzero coordinates of x. The *Hamming weight* $\mathrm{wt}(f)$ of the function $f \in \mathcal{F}_n$ is the Hamming weight of its vector of values. The *Hamming distance* $\mathrm{dist}(f, g)$ between Boolean functions f, g in n variables

is a cardinality of the set $\left\{x \in \mathbb{F}_2^n | f(x) \oplus g(x) = 1\right\}$. The *Walsh–Hadamard transform* (WHT) of the Boolean function f in n variables is an integer function $W_f : \mathbb{F}_2^n \to \mathbb{Z}$, defined as

$$W_f(y) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus \langle x, y \rangle}, \quad y \in \mathbb{F}_2^n.$$

A Boolean function f in an even number n of variables is said to be *bent* if

$$|W_f(y)| = 2^{n/2},$$

for all $y \in \mathbb{F}_2^n$. The set of bent functions in n variables is denoted by $\mathcal{B}_n$. From the definition of bent function f it follows that there exists such Boolean function $\widetilde{f} \in \mathcal{F}_n$ that for any $y \in \mathbb{F}_2^n$ we have

$$W_f(y) = (-1)^{\widetilde{f}(y)} 2^{n/2}.$$

The Boolean function $\widetilde{f}$ defined above is called the *dual* function of the bent function f. The duality of bent functions was introduced by Dillon [6].

Some known properties of dual functions [1]:

- Every dual function is a bent function;
- If $\widetilde{f}$ is dual to f and $\widetilde{\widetilde{f}}$ is dual to $\widetilde{f}$, then $\widetilde{\widetilde{f}} = f$;
- The mapping $f \to \widetilde{f}$ which acts on the set of bent functions, preserves the Hamming distance.

If bent function f coincides with its dual, that is $f = \widetilde{f}$, it is said to be *self-dual bent*. A bent function f which coincides with the negation of its dual, that is $f = \widetilde{f} \oplus 1$, is called an *anti-self-dual bent*. The set of (anti-)self-dual bent functions in n variables, according to [8], is denoted by $\mathrm{SB}^+(n)$ $\left(\mathrm{SB}^-(n)\right)$.

Denote, according to [10], the orthogonal group of index n over the field $\mathbb{F}_2$ as

$$\mathcal{O}_n = \left\{L \in GL(n, \mathbb{F}_2) \,|\, LL^T = I_n\right\},$$

where L^T denotes the transpose of L and I_n is an identity matrix of order n over the field $\mathbb{F}_2$.

3 Isometric mappings and automorphism groups

A mapping φ of the set of all Boolean functions in n variables to itself is called *isometric* if it preserves the Hamming distance between functions, that is

$$\mathrm{dist}(\varphi(f), \varphi(g)) = \mathrm{dist}(f, g),$$

for any $f, g \in \mathcal{F}_n$. The set of all isometric mappings of the set of all Boolean functions in n variables to itself is denoted by $\mathcal{I}_n$.

Example 1 Composition of an affine transform of coordinates and an affine shift, that is the mapping of the form

$$f(x) \longrightarrow f(Lx \oplus b) \oplus \langle c, x \rangle \oplus d, \tag{1}$$

where L is a $n \times n$ nonsingular binary matrix, $b, c \in \mathbb{F}_2^n$, $d \in \mathbb{F}_2$, is an element of $\mathcal{I}_n$.

The general form of isometric mappings of all Boolean functions in n variables to itself is

$$f(x) \longrightarrow f(\pi(x)) \oplus g(x),$$

where π is a permutation on the set $\mathbb{F}_2^n$ and $g \in \mathcal{F}_n$ [15]. The mapping of this form is denoted by $\varphi_{\pi,g} \in \mathcal{I}_n$.

Recall that a square matrix is called *monomial* (or *generalized permutation matrix*) if it has exactly one nonzero entry in each row and each column. There is an one-to-one correspondence between $\mathcal{I}_n$ and the set of monomial matrices of order $2^n \times 2^n$ with nonzero elements from the set $\{\pm 1\}$. Indeed, consider arbitrary mapping $\varphi_{\pi,g} \in \mathcal{I}_n$.

Then for any $f \in \mathcal{F}_n$ and its sign function

$$F = \left((-1)^{f(\mathbf{v}_0)}, (-1)^{f(\mathbf{v}_1)}, ..., (-1)^{f(\mathbf{v}_{2^n-1})}\right),$$

the sign function

$$F' = \left((-1)^{f'(\mathbf{v}_0)}, (-1)^{f'(\mathbf{v}_1)}, ..., (-1)^{f'(\mathbf{v}_{2^n-1})}\right),$$

of $f' = \varphi_{\pi,g}(f) \in \mathcal{F}_n$ can be expressed as $F' = AF$, where A is a $2^n \times 2^n$ monomial matrix, constructed by the permutation π and the function g:

$$\begin{array}{c} \\ i \end{array} \begin{array}{c} \begin{array}{c} j \end{array} \\ \begin{pmatrix} \vdots \\ 0 \\ \vdots \\ \dots\ 0\ \dots\ (-1)^{g(\mathbf{v}_{i-1})}\ \dots\ 0\ \dots \\ \vdots \\ 0 \\ \vdots \end{pmatrix} \end{array},$$

in which in the i-th row a nonzero element $(-1)^{g(\mathbf{v}_{i-1})}$ is in the j-th column, where $(j-1)$ is a number with binary representation $\pi(\mathbf{v}_{i-1})$. So the i-th component of the vector $F' = AF$ is equal to

$$(-1)^{f'(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))} \cdot (-1)^{g(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))\oplus g(\mathbf{v}_{i-1})},$$

where $i = 1, 2, ..., 2^n$, from which it follows that

$$f'(x) = f(\pi(x)) \oplus g(x), \quad x \in \mathbb{F}_2^n.$$

The *group of automorphisms* of a fixed subset $M \subseteq \mathcal{F}_n$ is the group of isometric mappings of the set of all Boolean functions in n variables to itself preserving the set M. It is denoted by $\mathrm{Aut}(M)$.

The group of automorphisms of the set of bent functions was completely characterized by Tokareva in 2010: it was proved that every isometric mapping of the set of all Boolean functions in an even number n of variables to itself that transforms bent functions to bent functions is a combination of an affine transform of coordinates and an affine shift [20], in other words, it is described by (1).

4 Sign functions of self-dual bent functions

A non-zero vector $v \in \mathbb{C}^n$ is called an *eigenvector* of a square $n \times n$ matrix A attached to the eigenvalue $\lambda \in \mathbb{C}$ if $Av = \lambda v$. A linear span of eigenvectors attached to the eigenvalue λ is called an *eigenspace* associated with λ.

Consider a linear mapping $\psi : \mathbb{C}^n \to \mathbb{C}^n$ represented by a $n \times n$ complex matrix A. A *kernel* of ψ is the set

$$\mathrm{Ker}\,(\psi) = \left\{x \in \mathbb{C}^n | Ax = \mathbf{0} \in \mathbb{C}^n\right\},$$

where $\mathbf{0}$ is a zero element of the space $\mathbb{C}^n$.

Let I_n be the identity matrix of size n and $H_n = H_1^{\otimes n}$ be the n-fold tensor product of the matrix H_1 with itself, where

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

It is known the Hadamard property of this matrix

$$H_n H_n^T = 2^n I_{2^n},$$

where H_n^T is transpose of H_n (it holds $H_n^T = H_n$ by symmetricity of H_n). Denote $\mathcal{H}_n = 2^{-n/2} H_n$, this matrix is symmetric and orthogonal.

Recall an orthogonal decomposition of $\mathbb{R}^{2^n}$ in eigenspaces of H_n from [2] (Lemma 5.2):

$$\mathbb{R}^{2^n} = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right) \oplus \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right),$$

where the symbol $\oplus$ denotes a direct sum of subspaces.

Since all rows of the matrix H_n correspond to sign functions of all linear functions (see [4] for instance), equivalently, a bent function can be defined as a function whose sign function, say F, satisfies $\mathcal{H}_n F \in \{\pm 1\}^{2^n}$. From the definition of self-duality it follows that sign function of any self-dual bent function is the eigenvector of $\mathcal{H}_n$ attached to the eigenvalue 1, that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n}) = \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$. The same holds for a sign function of any anti-self-dual bent function, which obviously is an eigenvector of $\mathcal{H}_n$ attached to the eigenvalue (-1), that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}) = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$.

It is known that

$$\dim\left(\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})\right) = \dim\left(\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})\right) = 2^{n-1},$$

where $\dim(V)$ is the dimension of the subspace $V \subseteq \mathbb{R}^{2^n}$. Moreover, from symmetricity of $\mathcal{H}_n$ it follows that

$$\left(\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})\right)^{\perp} = \mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$$

and

$$\left(\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})\right)^{\perp} = \mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}).$$

In [12] the following result was obtained:

Proposition 1 ([12], Theorem 2) *Let $n \geqslant 4$, then the linear span of sign functions of (anti-) self-dual bent functions in n variables has dimension 2^{n-1}.*

In other words, among sign functions of self-dual bent functions in n variables there exists a basis of the eigenspace of the matrix H_n attached to the eigenvalues 1, that is the subspace $\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$. Similarly, among sign functions of anti-self-dual bent functions in n variables there exists a basis of the eigenspace of the matrix H_n attached to the eigenvalues (-1), that is the subspace $\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$.

For $n = 2$ there are two self-dual bent functions, namely $x_1 x_2$ and $x_1 x_2 \oplus 1$, which have sign functions $(1, 1, 1, -1)$ and $(-1, -1, -1, 1)$ respectively. These sign functions are linearly dependent vectors in $\mathbb{R}^4$. The set $\mathrm{SB}^-(2)$ consists of functions $x_1 x_2 \oplus x_1 \oplus x_2$ and $x_1 x_2 \oplus x_1 \oplus x_2 \oplus 1$ with sign functions $(1, -1, -1, -1)$ and $(-1, 1, 1, 1)$ respectively. These sign functions are linearly dependent vectors in $\mathbb{R}^4$ as well.

5 Isometric mappings preserving self-duality

In [7] (Theorem 1) it was shown that the mapping

$$f(x) \longrightarrow f(L(x \oplus c)) \oplus \langle c, x \rangle \oplus d,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$, preserves self-duality of a bent function. It is obvious that every mapping of such form is an element of $\mathcal{I}_n$ with $\pi(x) = L(x \oplus c)$ and $g(x) = \langle c, x \rangle \oplus d$, $x \in \mathbb{F}_2^n$. The group which consists of mappings of such form is called an *extended orthogonal group* and denoted by $\overline{\mathcal{O}}_n$ [5, 7]. It holds $\overline{\mathcal{O}}_n \leqslant \mathrm{GL}(n+2, \mathbb{F}_2)$.

Assume that $n \geqslant 4$ is an even integer. In this section we generalize this result within isometric mappings from the set $\mathcal{I}_n$.

At first there is the question of how the sets of isometric mapping preserving self-duality and anti-self-duality or, in other words, automorphism groups of the sets $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ are connected.

Proposition 2 *For isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ with matrix A the following conditions are equivalent:*

1) *$\varphi_{\pi,g}$ preserves self-duality;*
2) *$\varphi_{\pi,g}$ preserves anti-self-duality;*
3) *$A\mathcal{H}_n = \mathcal{H}_n A$.*

Proof By Proposition 1 for $n \geqslant 4$ within the set $\mathrm{SB}^+(n)$ there exist a subset $\{f_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{SB}^+(n)$ with linearly independent sign functions $\{F_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{Ker}(\mathcal{H}_n - I_{2^n})$ and a subset $\{g_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{SB}^-(n)$ with linearly independent sign functions $\{G_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{Ker}(\mathcal{H}_n + I_{2^n})$.

Prove that from the first assertions of the Proposition the second one follows. Assume $\varphi_{\pi,g}$ preserves self-duality. Since the matrix A is a nonsingular one, the vectors $\{AF_i\}_{i=1}^{2^{n-1}}$ are also linearly independent sign functions of self-dual bent functions $\left\{\varphi_{\pi,g}(f_i)\right\}_{i=1}^{2^{n-1}} \subseteq \mathrm{SB}^+(n)$. Then for any sign function $R \in \mathrm{Ker}(\mathcal{H}_n + I_{2^n})$ of $r \in \mathrm{SB}^-(n)$ we have

$$\langle AR, AF_i \rangle = \left\langle A^T AR, F_i \right\rangle = \langle R, F_i \rangle = 0$$

for $i = 1, 2, ..., 2^{n-1}$, hence it holds $AR \in \mathrm{Ker}(\mathcal{H}_n + I_{2^n})$ and immediately $\varphi_{\pi,g}(r) \in \mathrm{SB}^-(n)$. That is, for every anti-self-dual bent function r its image $\varphi_{\pi,g}(r)$ is also an anti-self-dual bent function.

By using the same arguments one can show that from the second assertions the first one follows as well, and we can conclude that the first and the second ones are equivalent.

Now prove the equivalence of the first and the third assertions. If $A\mathcal{H}_n = \mathcal{H}_n A$, then for any sign function F of $f \in \mathrm{SB}^+(n)$ it holds

$$\mathcal{H}_n(AF) = A(\mathcal{H}_n F) = AF,$$

hence the mapping preserves self-duality.

Denote $B = \mathcal{H}_n A - A\mathcal{H}_n$ and assume that the mapping with matrix A preserves self-duality and, as proved above, anti-self-duality. In particular, for $i = 1, 2, ..., 2^{n-1}$ it holds

$$\mathcal{H}_n(AF_i) = AF_i$$

and

$$\mathcal{H}_n(AG_i) = -AG_i.$$

For $i = 1, 2, ..., 2^{n-1}$ we have:

$$(\mathcal{H}_n A - A\mathcal{H}_n) F_i = \mathcal{H}_n (AF_i) - A (\mathcal{H}_n F_i) = \mathcal{H}_n (AF_i) - AF_i = BF_i.$$

Then $BF_i = \mathbf{0} \in \mathbb{R}^{2^n}$ for every $i = 1, 2, ..., 2^{n-1}$. From the fact that the set $\{F_i\}_{i=1}^{2^{n-1}}$ forms a basis of the subspace $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$ it follows that all rows of the matrix B are vectors from the subspace $(\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n}))^{\perp} = \mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})$.

For $i = 1, 2, ..., 2^{n-1}$ we also have

$$(\mathcal{H}_n A - A\mathcal{H}_n) G_i = \mathcal{H}_n (AG_i) - A (\mathcal{H}_n G_i) = \mathcal{H}_n (AG_i) + AG_i = BG_i.$$

In this case $BG_i = \mathbf{0} \in \mathbb{R}^{2^n}$ for every $i = 1, 2, ..., 2^{n-1}$. Since the set $\{G_i\}_{i=1}^{2^{n-1}}$ forms a basis of the subspace $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})$ we can conclude that all rows of the matrix B are vectors from the subspace $(\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}))^{\perp} = \mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$.

Thus, we have proved that all rows of the matrix B lie in $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}) \cap \mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$ but the intersection of orthogonal subspaces consists only of the zero element of the space $\mathbb{R}^n$. Therefore the matrix B is zero matrix. □

Corollary 1 *It holds*

$$\mathrm{Aut}\left(\mathrm{SB}^{+}(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^{-}(n)\right).$$

The criterion (condition $A\mathcal{H}_n = \mathcal{H}_n A$) can be reformulated as follows: if $n \geqslant 4$ then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ belongs to $\mathrm{Aut}\left(\mathrm{SB}^{+}(n)\right)$ if and only if for any $x, y \in \mathbb{F}_2^n$ it holds

$$\langle \pi(x), y\rangle \oplus g(x) = \left\langle x, \pi^{-1}(y)\right\rangle \oplus g\left(\pi^{-1}(y)\right).$$

From Proposition 2 it follows that the problem of characterization of isometric mappings with considered properties is directly linked with the problem of enumerating all monomial matrices of order $2^n \times 2^n$ with elements from the set $\{0, \pm 1\}$, which commute with the matrix $\mathcal{H}_n$. The solution of this problem is given by the following

Theorem 1 *Isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ preserves (anti-)self-duality if and only if*

$$\pi(x) = L(x \oplus c), \ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \langle c, x\rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$.

Proof The opposite direction immediately comes from [7] (Theorem 1).

Assume that A is a matrix of the mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ preserving (anti-)self-duality. Let $T_{a,r}$ be a sign function of an affine function $l(x) = \langle a, x\rangle \oplus r$, where $a, x \in \mathbb{F}_2^n$, $r \in \mathbb{F}_2$. In other words $T_{a,r}$ is equal to some row (column) of the matrix H_n if $r = 0$ or $(-H_n)$ in the case $r = 1$. From Proposition 2 it follows that $A\mathcal{H}_n = \mathcal{H}_n A$ hence

$$\mathcal{H}_n\left(AT_{a,r}\right) = A\left(\mathcal{H}_n T_{a,r}\right) = 2^{n/2}\sigma \cdot Ae_k = 2^{n/2}\sigma' \cdot e_{k'},$$

where $k, k' \in \{1, 2, ..., 2^n\}$, $\sigma, \sigma' \in \{\pm 1\}$. Then

$$AT_{a,r} = 2^{n/2}\sigma' \cdot \mathcal{H}_n e_{k'} = T_{a',r'}$$

for some $a' \in \mathbb{F}_2^n$, $r' \in \mathbb{F}_2$.

Thus, the considered mapping transforms the set of all affine functions in n variables to itself hence it has form

$$f(x) \longrightarrow f(Lx \oplus b) \oplus \langle c, x\rangle \oplus d,$$

where L is a $n \times n$ nonsingular binary matrix, $b, c \in \mathbb{F}_2^n$, $d \in \mathbb{F}_2$, see [14], for example.

Now consider the relation $AH_n = H_nA$ in details. Recall that

$$H_n = \begin{pmatrix} (-1)^{\langle \mathbf{v}_0, \mathbf{v}_0 \rangle} & (-1)^{\langle \mathbf{v}_0, \mathbf{v}_1 \rangle} & \dots & (-1)^{\langle \mathbf{v}_0, \mathbf{v}_{2^n-1} \rangle} \\ (-1)^{\langle \mathbf{v}_1, \mathbf{v}_0 \rangle} & (-1)^{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} & \dots & (-1)^{\langle \mathbf{v}_1, \mathbf{v}_{2^n-1} \rangle} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{\langle \mathbf{v}_{2^n-1}, \mathbf{v}_0 \rangle} & (-1)^{\langle \mathbf{v}_{2^n-1}, \mathbf{v}_1 \rangle} & \dots & (-1)^{\langle \mathbf{v}_{2^n-1}, \mathbf{v}_{2^n-1} \rangle} \end{pmatrix}.$$

The i-th row of the matrix A has the following form

$$\begin{matrix} & & & & j & & & \\ \Big(& 0 & \dots & 0 & (-1)^{\langle c, \mathbf{v}_{i-1} \rangle \oplus d} & 0 & \dots & 0 \Big), \end{matrix}$$

where the nonzero element $(-1)^{\langle c, \mathbf{v}_{i-1} \rangle \oplus d}$ is in the j-th column, where $(j-1)$ is a number with binary representation $(L\mathbf{v}_{i-1} \oplus b)$.

Fix arbitrary $i, j \in \{0, 1, ..., 2^n - 1\}$. Write explicitly

$$(AH_n)_{i+1,j+1} = (-1)^{\langle c, \mathbf{v}_i \rangle \oplus \langle L\mathbf{v}_i \oplus b, \mathbf{v}_j \rangle \oplus d}.$$

The j-th column of the matrix A has the following form

$$\begin{matrix} & & & & & & & \\ i & \Bigg(& 0 & & & & & \\ & & \vdots & & & & & \\ & & 0 & & & & & \\ & & (-1)^{\langle c, L^{-1}(\mathbf{v}_{j-1} \oplus b) \rangle \oplus d} & & & & & \\ & & 0 & & & & & \\ & & \vdots & & & & & \\ & & 0 & & & & & \Bigg) \end{matrix}, ,$$

where the nonzero element $(-1)^{\langle c, L^{-1}(\mathbf{v}_{j-1} \oplus b) \rangle \oplus d}$ is in the i-th row, where $(i-1)$ is a number with binary representation $L^{-1}\left(\mathbf{v}_{j-1} \oplus b\right)$.

Then it clear that

$$(H_nA)_{i+1,j+1} = (-1)^{\langle \mathbf{v}_i, L^{-1}(\mathbf{v}_j \oplus b) \rangle \oplus \langle c, L^{-1}(\mathbf{v}_j \oplus b) \rangle \oplus d}.$$

Since $AH_n = H_nA$ implies $(AH_n)_{i+1,j+1} = (H_nA)_{i+1,j+1}$ for any $i, j \in \{0, 1, ..., 2^n - 1\}$, the following relation must hold

$$(-1)^{\langle c, \mathbf{v}_i \rangle \oplus \langle L\mathbf{v}_i \oplus b, \mathbf{v}_j \rangle \oplus d} = (-1)^{\langle \mathbf{v}_i, L^{-1}(\mathbf{v}_j \oplus b) \rangle \oplus \langle c, L^{-1}(\mathbf{v}_j \oplus b) \rangle \oplus d},$$

or, equivalently,

$$\langle c, x \rangle \oplus \langle Lx \oplus b, y \rangle \oplus d = \left\langle x, L^{-1}(y \oplus b) \right\rangle \oplus \left\langle c, L^{-1}(y \oplus b) \right\rangle \oplus d \tag{2}$$

for any $x, y \in \mathbb{F}_2^n$.

Put zero vector $y \in \mathbb{F}_2^n$ in (2). Then

$$\langle c, x \rangle = \left\langle x, L^{-1}b \right\rangle \oplus \left\langle c, L^{-1}b \right\rangle,$$

$$\left\langle x, L^{-1}b \oplus c \right\rangle = \left\langle c, L^{-1}b \right\rangle$$

for any $x \in \mathbb{F}_2^n$. Then

$$\begin{cases} L^{-1}b \oplus c = 0, \\ \langle c, L^{-1}b \rangle = 0, \end{cases}$$

$$\begin{cases} b = Lc, \\ \mathrm{wt}\,(c) \text{ is even.} \end{cases} \tag{3}$$

Return to (2) and take (3) into account:

$$\langle c, x \rangle \oplus \langle Lx \oplus Lc, y \rangle = \left\langle x, L^{-1}\left(y \oplus Lc\right) \right\rangle \oplus \left\langle c, L^{-1}\left(y \oplus Lc\right) \right\rangle,$$

$$\langle c, x \rangle \oplus \langle Lx, y \rangle \oplus \langle Lc, y \rangle = \left\langle x, L^{-1}y \right\rangle \oplus \langle x, c \rangle \oplus \left\langle c, L^{-1}y \right\rangle \oplus \langle c, c \rangle,$$

$$\langle Lx, y \rangle \oplus \langle Lc, y \rangle = \left\langle x, L^{-1}y \right\rangle \oplus \left\langle c, L^{-1}y \right\rangle,$$

$$\langle L\left(x \oplus c\right), y \rangle = \left\langle \left(L^{-1}\right)^T \left(x \oplus c\right), y \right\rangle.$$

for any $x, y \in \mathbb{F}_2^n$. In this case

$$L\left(x \oplus c\right) = \left(L^{-1}\right)^T \left(x \oplus c\right)$$

for any $x \in \mathbb{F}_2^n$ that is

$$L(z) = \left(L^{-1}\right)^T (z)$$

for any $z \in \mathbb{F}_2^n$. It holds if and only if

$$L = \left(L^{-1}\right)^T. \tag{4}$$

Thus, combining (3) and (4) we obtain

$$\begin{cases} L^{-1} = L^T, \\ b = Lc, \\ \mathrm{wt}\,(c) \text{ is even.} \end{cases}$$

□

Corollary 2 *It holds*

$$\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \overline{\mathcal{O}}_n.$$

It can be concluded that from Proposition 2 and Theorem 1 it follows that the group of automorphisms of the set of (anti-)self-dual bent functions coincides with the extended orthogonal group, that is

$$\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right) = \overline{\mathcal{O}}_n.$$

5.1 Sets of (anti-)self-dual bent function in two variables

The case $n = 2$ is out of the ordinary, because, in particular, Propositions 1 and 2 do not hold. Indeed, consider isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_2$ with the followong matrix:

$$A = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

It transforms sign function $(1, 1, 1, -1)$ of self-dual bent function $f(x_1, x_2) = x_1x_2$ to its negation $(-1, -1, -1, 1)$ and sign function $(1, -1, -1, -1)$ of anti-self-dual bent function $f(x_1, x_2) = x_1x_2 \oplus x_1 \oplus x_2$ to itself, that is this isometric mapping preserves both self-duality and anti-self-duality. But we have

$$A\mathcal{H}_n = \begin{pmatrix} -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}, \quad \mathcal{H}_n A = \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1. \end{pmatrix},$$

and $A\mathcal{H}_n \neq \mathcal{H}_n A$.

Consider another isometric mapping $\varphi_{\pi',g'} \in \mathcal{I}_2$ with the followong matrix:

$$A' = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}.$$

It transforms sign function $(1, 1, 1, -1)$ of the self-dual bent function $f(x_1, x_2) = x_1x_2$ to itself but sign function $(1, -1, -1, -1)$ of the anti-self-dual bent function $f(x_1, x_2) = x_1x_2 \oplus x_1 \oplus x_2$ it transforms to sign function $(-1, 1, -1, -1)$ of bent function $f(x_1, x_2) = x_1x_2 \oplus x_2 \oplus 1$ which is neither self-dual nor anti-self-dual, that is this isometric mapping preserves self-duality but does not preserve anti-self-duality.

6 Isometric bijections between self-dual and anti-self-dual bent functions

It is known [2] (Theorems 5.1, 5.3) that there exists a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$, based on the decomposition of sign functions of (anti-)self-dual bent functions. Also note that from the existence of such bijection it follows that $\left|\mathrm{SB}^+(n)\right| = \left|\mathrm{SB}^-(n)\right|$.

Namely, let $(Y, Z) \in \{\pm 1\}^{2^n}$, where $Y, Z \in \{\pm 1\}^{2^{n-1}}$, be a sign function for some $f \in \mathrm{SB}^+(n)$. Then a vector $(Z, -Y) \in \{\pm 1\}^{2^n}$ is a sign function for some function from $\mathrm{SB}^-(n)$. In terms of isometric mappings the mentioned transform can be represented as

$$f(x) \longrightarrow f(x \oplus c) \oplus \langle c, x \rangle,$$

where $c = (1, 0, 0, ..., 0) \in \mathbb{F}_2^n$.

In paper [8] it was mentioned that the more general form of this mapping

$$f(x) \longrightarrow f(x \oplus c) \oplus \langle c, x \rangle,$$

where $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$. It is obvious that every mapping of such form is an element of $\mathcal{I}_n$.

Assume that $n \geqslant 4$ is an even integer. In this section we generalize these results within isometric mappings from the set $\mathcal{I}_n$.

Proposition 3 *Isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ with matrix A is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if $A\mathcal{H}_n = -\mathcal{H}_n A$.*

Proof If $\mathcal{H}_n A = -A\mathcal{H}_n$, then for any sign functions F, R of $f \in \mathrm{SB}^+(n)$ and $r \in \mathrm{SB}^-(n)$ respectively it holds

$$\mathcal{H}_n(AF) = -A(\mathcal{H}_n F) = -AF,$$

$$\mathcal{H}_n (AR) = -A (\mathcal{H}_n R) = AR,$$

hence the mapping is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$.

Take $\{f_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{SB}^+(n)$ with linearly independent sign functions $\{F_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$ and $\{g_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{SB}^-(n)$ with linearly independent sign functions $\{G_i\}_{i=1}^{2^{n-1}} \subseteq \mathrm{Ker}\,(H_n + I_{2^n})$ from the proof of the Proposition 2. Denote $B = \mathcal{H}_n A + A\mathcal{H}_n$ and assume that the mapping with matrix A is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$. In particular, for $i = 1, 2, ..., 2^{n-1}$ it holds

$$\mathcal{H}_n (AF_i) = -AF_i$$

and

$$\mathcal{H}_n (AG_i) = AG_i.$$

For $i = 1, 2, ..., 2^{n-1}$ we have

$$(\mathcal{H}_n A + A\mathcal{H}_n) F_i = \mathcal{H}_n (AF_i) + A (\mathcal{H}_n F_i) = \mathcal{H}_n (AF_i) + AF_i = BF_i.$$

Then $BF_i = \mathbf{0} \in \mathbb{R}^{2^n}$ for every $i = 1, 2, ..., 2^{n-1}$. Since the set $\{F_i\}_{i=1}^{2^{n-1}}$ forms a basis of the subspace $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$, it can be deduced that all rows of the matrix B are vectors from the subspace $(\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n}))^{\perp} = \mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})$.

For $i = 1, 2, ..., 2^{n-1}$ we also have:

$$(\mathcal{H}_n A + A\mathcal{H}_n) G_i = \mathcal{H}_n (AF_i) + A (\mathcal{H}_n G_i) = \mathcal{H}_n (AG_i) - AG_i = BG_i.$$

In this case $BG_i = \mathbf{0} \in \mathbb{R}^{2^n}$ for every $i = 1, 2, ..., 2^{n-1}$. Since the set $\{G_i\}_{i=1}^{2^{n-1}}$ forms a basis of the subspace $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})$ we can conclude that all rows of the matrix B are vectors from the subspace $(\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}))^{\perp} = \mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$.

Thus, we have proved that all rows of the matrix B lie in $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}) \cap \mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$ but the intersection of orthogonal subspaces consists only of the zero element of the space $\mathbb{R}^n$. Therefore the matrix B is zero matrix. □

The criterion (condition $A\mathcal{H}_n = -\mathcal{H}_n A$) can be reformulated as follows: if $n \geqslant 4$ then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ is a bijection between the sets $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if for any $x, y \in \mathbb{F}_2^n$ it holds

$$\langle \pi(x), y \rangle \oplus g(x) = \left\langle x, \pi^{-1}(y) \right\rangle \oplus g\left(\pi^{-1}(y)\right) \oplus 1.$$

Theorem 2 *Isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if*

$$\pi(x) = L (x \oplus c), \ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, $d \in \mathbb{F}_2$.

Proof Let $f \in \mathrm{SB}^+(n) \cup \mathrm{SB}^-(n)$ that is $\widetilde{f} = f \oplus \varepsilon$ for some $\varepsilon \in \mathbb{F}_2$. Consider a function $g(x) = f (L (x \oplus c)) \oplus \langle c, x \rangle \oplus d$, where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, $d \in \mathbb{F}_2$. Its

Walsh-Hadamard transform is

$$\begin{aligned} W_g(y) &= \sum_{x\in\mathbb{F}_2^n}(-1)^{\langle x,y\rangle\oplus g(x)} = \sum_{x\in\mathbb{F}_2^n}(-1)^{\langle x,y\rangle\oplus f(L(x\oplus c))\oplus\langle c,x\rangle\oplus d} \\ &= (-1)^d\sum_{x\in\mathbb{F}_2^n}(-1)^{\langle x,y\oplus c\rangle\oplus f(L(x\oplus c))} = (-1)^d\sum_{z\in\mathbb{F}_2^n}(-1)^{\langle L^{-1}z\oplus c,y\oplus c\rangle\oplus f(z)} \\ &= (-1)^{d\oplus\langle c,y\rangle\oplus\langle c,c\rangle}\sum_{z\in\mathbb{F}_2^n}(-1)^{\langle z,L(y\oplus c)\rangle\oplus f(z)} \\ &= (-1)^{d\oplus\langle c,y\rangle\oplus 1}2^{n/2}(-1)^{\widetilde{f}(L(y\oplus c))} = 2^{n/2}(-1)^{f(L(y\oplus c))\oplus\langle c,y\rangle\oplus d\oplus\varepsilon\oplus 1} \\ &= 2^{n/2}(-1)^{g(y)\oplus\varepsilon\oplus 1} = 2^{n/2}(-1)^{\widetilde{g}(y)}, \end{aligned}$$

hence $\widetilde{g}(y) = g(y)\oplus\varepsilon\oplus 1$ for any $y\in\mathbb{F}_2^n$. The opposite direction has been proved.

By using the same arguments as in the proof of the Theorem 1 it can be deduced that the considered isometric mapping preserves affinity of a Boolean function and therefore has form

$$f(x)\longrightarrow f\left(Lx\oplus b\right)\oplus\langle c,x\rangle\oplus d,$$

where L is a $n\times n$ nonsingular binary matrix, $b, c\in\mathbb{F}_2^n, d\in\mathbb{F}_2$.

From Proposition 3 it follows that $AH_n = -H_nA$. Recall from the proof of the Theorem 1 that

$$(AH_n)_{i+1,j+1} = (-1)^{\langle c,\mathbf{v}_i\rangle\oplus\langle L\mathbf{v}_i\oplus b,\mathbf{v}_j\rangle\oplus d},$$

$$(H_nA)_{i+1,j+1} = (-1)^{\langle\mathbf{v}_i,L^{-1}(\mathbf{v}_j\oplus b)\rangle\oplus\langle c,L^{-1}(\mathbf{v}_j\oplus b)\rangle\oplus d},$$

for any $i, j\in\{0, 1, ..., 2^n-1\}$.

Since $AH_n = -H_nA$ implies $(AH_n)_{i+1,j+1} = -(H_nA)_{i+1,j+1}$ for any $i, j\in\{0, 1, ..., 2^n-1\}$, the following relation must hold

$$(-1)^{\langle c,\mathbf{v}_i\rangle\oplus\langle L\mathbf{v}_i\oplus b,\mathbf{v}_j\rangle\oplus d} = (-1)^{\langle\mathbf{v}_i,L^{-1}(\mathbf{v}_j\oplus b)\rangle\oplus\langle c,L^{-1}(\mathbf{v}_j\oplus b)\rangle\oplus d\oplus 1},$$

or, equivalently,

$$\langle c,x\rangle\oplus\langle Lx\oplus b,y\rangle\oplus d = \left\langle x,L^{-1}(y\oplus b)\right\rangle\oplus\left\langle c,L^{-1}(y\oplus b)\right\rangle\oplus d\oplus 1 \tag{5}$$

for any $x, y\in\mathbb{F}_2^n$.

Put zero vector $y\in\mathbb{F}_2^n$ in (5). Then

$$\langle c,x\rangle = \left\langle x,L^{-1}b\right\rangle\oplus\left\langle c,L^{-1}b\right\rangle\oplus 1,$$

$$\left\langle x,L^{-1}b\oplus c\right\rangle = \left\langle c,L^{-1}b\right\rangle\oplus 1$$

for any $x\in\mathbb{F}_2^n$. Then

$$\begin{cases} L^{-1}b\oplus c = 0, \\ \langle c,L^{-1}b\rangle = 1, \end{cases}$$

$$\begin{cases} b = Lc, \\ \mathrm{wt}(c)\text{ is odd.} \end{cases} \tag{6}$$

Return to (5) and take (6) into account:

$$\langle c,x\rangle\oplus\langle Lx\oplus Lc,y\rangle = \left\langle x,L^{-1}(y\oplus Lc)\right\rangle\oplus\left\langle c,L^{-1}(y\oplus Lc)\right\rangle\oplus 1,$$

$$\langle c,x\rangle\oplus\langle Lx,y\rangle\oplus\langle Lc,y\rangle = \left\langle x,L^{-1}y\right\rangle\oplus\langle x,c\rangle\oplus\left\langle c,L^{-1}y\right\rangle\oplus\langle c,c\rangle\oplus 1,$$

$$\langle Lx, y\rangle \oplus \langle Lc, y\rangle = \left\langle x, L^{-1}y\right\rangle \oplus \left\langle c, L^{-1}y\right\rangle,$$

$$\langle L(x \oplus c), y\rangle = \left\langle \left(L^{-1}\right)^{T}(x \oplus c), y\right\rangle$$

for any $x, y \in \mathbb{F}_2^n$. It holds if and only if

$$L = \left(L^{-1}\right)^{T}. \tag{7}$$

Thus, combining (6) and (7) we obtain

$$\begin{cases} L^{-1} = L^{T}, \\ b = Lc, \\ \text{wt}(c) \text{ is odd.} \end{cases}$$

□

7 Isometric mappings and the Rayleigh quotient of the Sylvester Hadamard matrix

In this section isometric mappings from the set $\mathcal{I}_n$, which preserve and change the sign of the Rayleigh quotient (Rayleigh ratio) of the Sylvester Hadamard matrix defined for every Boolean function in n variables, are studied.

7.1 Definition and characterization

In [2] the *Rayleigh quotient* S_f of a Boolean function $f \in \mathcal{F}_n$ was defined as

$$S_f = \sum_{x,y\in\mathbb{F}_2^n} (-1)^{f(x)\oplus f(y)\oplus\langle x,y\rangle} = \sum_{y\in\mathbb{F}_2^n} (-1)^{f(y)} W_f(y).$$

For any $f \in \mathcal{B}_n$ the *normalized Rayleigh quotient* N_f is a number

$$N_f = \sum_{x\in\mathbb{F}_2^n} (-1)^{f(x)\oplus\widetilde{f}(x)} = 2^{-n/2} S_f.$$

In [2] (Theorem 3.1) it was proved that for any $f \in \mathcal{F}_n$ the absolute value of S_f is at most $2^{3n/2}$ with equality if and only if f is self-dual $\left(+2^{3n/2}\right)$ and anti-self-dual $\left(-2^{3n/2}\right)$ bent function.

In the article [5] the operations on Boolean functions that preserve bentness and the Rayleigh quotient were given. Namely, it was proved that for any $f \in \mathcal{B}_n$, $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $d \in \mathbb{F}_2$ the functions $g, h \in \mathcal{B}_n$ defined as $g(x) = f(Lx) \oplus d$ and $h(x) = f(x \oplus c) \oplus \langle c, x\rangle$ provide $N_g = N_f$ and $N_h = (-1)^{\langle c,c\rangle} N_f$.

One can notice that the mentioned operations are isometric mappings from $\mathcal{I}_n$.

Assume that $n \geqslant 4$ is an even integer. In the following subsections we generalize these results within isometric mappings from the set $\mathcal{I}_n$.

7.2 Isometric mappings preserving the Rayleigh quotient

Theorem 3 *Isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ preserves the Rayleigh quotient if and only if it preserves self-duality.*

Proof For straight direction it is enough to mention that $S_f = +2^{3n/2}$ if and only if $f \in \mathrm{SB}^+(n)$ ([2], Theorem 3.1).

Assume that the mapping $\varphi_{\pi,g}$ preserves self-duality. Let A be its matrix. Then by Proposition 2 we have $AH_n = H_nA$. Take arbitrary $f \in \mathcal{F}_n$ and rewrite the Rayleigh quotient in the following form:

$$S_f = \sum_{x,y \in \mathbb{F}_2^n} (-1)^{f(x) \oplus f(y) \oplus \langle x,y \rangle} = \langle F, H_n F \rangle ,$$

where F is a sign function of f. The mapping preserves the Rayleigh quotient if

$$S_{\varphi_{\pi,g}(f)} = \langle AF, H_n (AF) \rangle = \langle F, H_n F \rangle = S_f .$$

Consider

$$\langle AF, H_n (AF) \rangle = \langle AF, A (H_n F) \rangle = \left\langle A^T AF, H_n F \right\rangle = \langle F, H_n F \rangle ,$$

therefore $\varphi_{\pi,g}$ preserves the Rayleigh quotient. □

Corollary 3 *Isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ preserves the Rayleigh quotient if and only if*

$$\pi(x) = L (x \oplus c) , \ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$.

7.3 Isometric mappings changing the sign of the Rayleigh quotient

Theorem 4 *Isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ changes the sign of the Rayleigh quotient if and only if it is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$.*

Proof For straight direction it is enough to mention that $S_f = +2^{3n/2}$ if and only if $f \in \mathrm{SB}^+(n)$ and $S_f = -2^{3n/2}$ if and only if $f \in \mathrm{SB}^-(n)$ ([2], Theorem 3.1).

Assume that the mapping $\varphi_{\pi,g}$ is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$. Let A be its matrix. Then by Proposition 3 we have $AH_n + H_nA = 0$. Take arbitrary $f \in \mathcal{F}_n$ and rewrite the Rayleigh quotient in the following form:

$$S_f = \sum_{x,y \in \mathbb{F}_2^n} (-1)^{f(x) \oplus f(y) \oplus \langle x,y \rangle} = \langle F, H_n F \rangle ,$$

where F is a sign function of f. The mapping changes the sign of the Rayleight quotient if

$$S_{\varphi_{\pi,g}(f)} = \langle AF, H_n (AF) \rangle = - \langle F, H_n F \rangle = -S_f .$$

Consider

$$\langle AF, H_n (AF) \rangle = \langle AF, -A (H_n F) \rangle = - \left\langle A^T AF, H_n F \right\rangle = - \langle F, H_n F \rangle ,$$

therefore $\varphi_{\pi,g}$ changes the sign of the Rayleigh quotient. □

Corollary 4 *Isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ changes the sign of the Rayleigh quotient if and only if*

$$\pi(x) = L (x \oplus c) , \ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ *is odd,* $d \in \mathbb{F}_2$.

From Theorems 3 and 4 it follows

Corollary 5 *Isometric mapping* $\varphi_{\pi,g} \in \mathcal{I}_n$, *which preserves the Rayleigh quotient or changes the sign of the Rayleigh quotient, also preserves bentness.*

7.4 Isometric mappings preserving the Hamming distance between bent function and its dual

The Rayleigh quotient characterizes the Hamming distance between a bent-function and its dual. Indeed, let $f \in \mathcal{B}_n$, then

$$\mathrm{dist}\left(f, \widetilde{f}\right) = 2^{n-1} - \frac{1}{2^{n/2+1}} S_f = 2^{n-1} - \frac{1}{2} N_f.$$

Theorem 5 *Isometric mapping* $\varphi_{\pi,g} \in \mathcal{I}_n$ *preserves bentness and the Hamming distance between any bent function in n variables and its dual if and only if it preserves (anti-)self-duality.*

Proof If $\varphi_{\pi,g}$ preserves the Hamming distance between any bent function in n variables and its dual then it preserves (anti-)self-duality.

If $\varphi_{\pi,g}$ preserves (anti-)self-duality then by Theorem 3 it preserves the Rayleigh quotient and from Theorem 1 it follows that this mapping preserves bentness. The characterization of the Hamming distance between a bent function and its dual in terms of the Rayleigh quotient yields the result. □

The form of such mappings is described by Theorem 1.

8 Summary

In this section we summarize and group results from the paper.

Assume that $n \geqslant 4$ is an even integer.

Let $\varphi_{\pi,g}$ be an isometric mapping of the set of all Boolean functions in n variables to itself with matrix A, namely

$$\varphi_{\pi,g} : f(x) \longrightarrow f(\pi(x)) \oplus g(x),$$

where π is a permutation in $\mathbb{F}_2^n$ and $g \in \mathcal{F}_n$. The matrix A is the following

$$\begin{array}{c} \\ i \end{array} \begin{array}{c} \begin{array}{c} j \end{array} \\ \begin{pmatrix} \vdots \\ 0 \\ \vdots \\ \ldots\ 0\ \ldots\ (-1)^{g(\mathbf{v}_{i-1})}\ \ldots\ 0\ \ldots \\ \vdots \\ 0 \\ \vdots \end{pmatrix} \end{array},$$

in which in the i-th row a nonzero element $(-1)^{g(\mathbf{v}_{i-1})}$ is in the j-th column, where $(j-1)$ is a number with binary representation $\pi(\mathbf{v}_{i-1})$.

Theorem 6 *The following conditions are equivalent:*

1) $\varphi_{\pi,g}$ *preserves self-duality;*
2) $\varphi_{\pi,g}$ *preserves anti-self-duality;*
3) $\varphi_{\pi,g}$ *preserves the Rayleigh quotient of every Boolean function;*
4) $\varphi_{\pi,g}$ *preserves bentness and the Hamming distance between any bent function and its dual;*
5) $\pi(x) = L(x \oplus c)$, $g(x) = \langle c, x\rangle \oplus d$, *where* $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ *is even,* $d \in \mathbb{F}_2$;
6) $A\mathcal{H}_n = \mathcal{H}_n A$.

Theorem 7 *The following conditions are equivalent:*

1) $\varphi_{\pi,g}$ *is a bijection between* $\mathrm{SB}^+(n)$ *and* $\mathrm{SB}^-(n)$;
2) $\varphi_{\pi,g}$ *changes sign of the Rayleigh quotient of every Boolean function;*
3) $\pi(x) = L(x \oplus c)$, $g(x) = \langle c, x\rangle \oplus d$, *where* $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ *is odd,* $d \in \mathbb{F}_2$;
4) $A\mathcal{H}_n = -\mathcal{H}_n A$.

Recall that the extended orthogonal group $\overline{\mathcal{O}}_n$ consists of mappings of all Boolean functions in n variables to itself which have form

$$f(x) \longrightarrow f(L(x \oplus c)) \oplus \langle c, x\rangle \oplus d,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$.

The group of automorphisms of (anti-)self-dual bent functions is characterized by the following

Theorem 8 *It holds*

$$\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right) = \overline{\mathcal{O}}_n.$$

From the obtained results it follows that an approach to equivalence of self-dual bent functions in $n \geqslant 4$ variables based on the restricted form of affine equivalence proposed in articles [2, 7] is the most general within isometric mappings of the set of all Boolean functions in n variables to itself.

9 Conclusion

In current paper isometric mappings of all Boolean functions in $n \geqslant 4$ variables to itself preserving self-duality and anti-self-duality of a Boolean bent function were completely studied. The obtained results were used to determine isometric mappings preserving the Rayleigh quotient of a Boolean function and isometric mappings preserving bentness and the Hamming distance between any bent function and its dual. The group of automorphisms of the set of (anti-)self-dual bent functions is obtained.

An interesting open problem is to characterize isometric mappings preserving self-duality which are not necessarily isometric mappings of the set of all Boolean functions in n variables.

References

1. Carlet, C.: Boolean functions for cryptography and error correcting code. In: Crama, Y., Hammer, P.L. (eds.) Boolean models and methods in mathematics, computer science, and engineering, pp. 257-397. Cambridge University Press, Cambridge (2010)
2. Carlet, C., Danielson, L.E., Parker, M.G., Solé, P.: Self-dual bent functions. Int. J. Inform. Coding Theory **1**, 384–399 (2010)
3. Carlet, C., Mesnager, S.: Four decades of research on bent functions. Journal Des. Codes Cryptogr. Springer **78**(1), 5–50 (2016)
4. Cusick, T.W., Stănică P.: Cryptographic Boolean functions and applications, p. 288. Acad. Press, London (2017)
5. Danielsen, L.E., Parker, M.G., Solé, P.: The Rayleigh quotient of bent functions, Springer Lect. Notes in Comp. Sci. 5921, pp. 418–432. Springer, Berlin (2009)
6. Dillon, J.: Elementary Hadamard Difference Sets, PhD. dissertation. College Park, Univ Maryland (1974)
7. Feulner, T., Sok, L., Solé, P., Wassermann, A.: Towards the classification of self-dual bent functions in eight variables. Des. Codes Cryptogr. **68**(1), 395–406 (2013)
8. Hou, X.-D.: Classification of self dual quadratic bent functions. Des Codes Cryptogr. **63**(2), 183–198 (2012)
9. Hyun, J.Y., Lee, H., Lee, Y.: MacWilliams duality and Gleason-type theorem on self-dual bent functions. Des Codes Cryptogr. **63**(3), 295–304 (2012)
10. Janusz, G.J.: Parametrization of self-dual codes by orthogonal matrices. Finite Fields Appl. **13**(3), 450–491 (2007)
11. Kutsenko, A.V.: The Hamming distance spectrum between self-dual Maiorana–McFarland bent functions. J. Appl. Ind. Math. **12**(1), 112–125 (2018)
12. Kutsenko, A.: Metrical properties of self-dual bent functions. Des. Codes Cryptogr. **88**(1), 201–222 (2020)
13. Luo, G., Cao, X., Mesnager, S.: Several new classes of self-dual bent functions derived from involutions. Cryptogr. Commun. **11**(6), 1261–1273 (2019)
14. MacWilliams, F.J., Sloane, N.J.A.: The theory of error correcting codes. North-Holland, Amsterdam (1977)
15. Markov A.A.: On transformations without error propagation. In: Selected works, Vol. II: Theory of algorithms and constructive mathematics. Mathematical Logic. Informatics and Related Topics, p. 70–93, MTsNMO, Moscow [Russian] (2003)
16. Mesnager, S.: Several new infinite families of bent functions and their duals. IEEE Trans. Inf. Theory **60**(7), 4397–4407 (2014)
17. Mesnager, S.: Bent functions: Fundamentals and results, p. 544. Springer, Berlin (2016)
18. Rothaus, O.S.: On bent functions. J. Combin. Theory. Ser. A **20**(3), 300–305 (1976)
19. Sok, L., Shi, M., Solé, P.: Classification and Construction of quaternary self-dual bent functions. Cryptogr. Commun. **10**(2), 277–289 (2018)
20. Tokareva, N.: The group of automorphisms of the set of bent functions. Discret. Math. Appl. **20**(5), 655–664 (2010)
21. Tokareva, N.: Bent functions: Results and applications to cryptography, 230 p., Acad. Press Elsevier (2015)

Properties of the secret gamma in stream ciphers

1st Tatiana Bonich
Novosibirsk State University
Laboratory of Cryptography
JetBrains Research
Novosibirsk, Russia
t.bonich@g.nsu.ru

2nd Matvey Panferov
Novosibirsk State University
Laboratory of Cryptography
JetBrains Research
Novosibirsk, Russia
m.panferov@g.nsu.ru

3rd Natalia Tokareva
Sobolev Institute of Mathematics
Novosibirsk, Russia
tokareva@math.nsc.ru

***Abstract*—It is well known that every stream cipher is based on a good pseudorandom generator. For cryptographic purposes we are interested in generation of pseudorandom sequences of the maximal possible period. A feedback register is one of the most known cryptographic primitives that is used in construction of stream generators. In this paper we analyze periodic properties of pseudorandom sequences produced by filter and combiner generators (two known schemes of stream generators based on feedback registers). We determine which nonlinear functions in these schemes lead to pseudorandom sequences of the not maximal possible length. We call such functions unsuitable and count the exact number of them for an arbitrary n.**

***Index Terms*—stream cipher, filter generator, combiner generator, gamma, Boolean function**

I. Introduction

Symmetric ciphers usually are divided into block and stream ones. Stream ciphers are considered as more fast but not as secure as block ciphers. One of the most known cryptographic primitives that is used for stream ciphers construction is a feedback register. There are attacks [6] and defenses [5] on such ciphers.

Remember that a *feedback shift register (FSR)* contains two parts: a binary block $x = (x_{n-1}, \dots x_0)$ of length n and a feedback function $f : (x_{n-1}, \dots, x_0) \to \{0, 1\}$, where f is a Boolean function in n variables. First, we fill the block x with concrete values of bits; together they form the *initial state* of the register. For functioning of the FSR the time is considered to be discrete, i. e. it is divided into clock cycles. On each clock cycle, the value of $f(x)$ is calculated first, then the state $x = (x_{n-1}, \dots, x_1, x_0)$ of the register will be changed to the state $x' = (x_{n-2}, \dots x_0, f(x))$ while the bit x_{n-1} will be written as the first bit of the generated sequence *gamma*.

The properties of gamma generated by FSR are well studied in case when f is a linear function. If f is nonlinear, [7], then there are too many open questions with properties of gamma that all are connected to analysis of nonlinear recurrent sequences, [4] and [3]. That is why in cryptography some nonlinear *combinations* of linear FSRs are considered, for instance —- filter and combining models of stream generators based on LFSR, [1].

The work was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by Russian Foundation for Basic Research (project no. 18-07-01394) and Laboratory of Cryptography JetBrains Research.

In this paper we analyze pseudorandom sequences produced by filter and combiner generators. Namely, we study which nonlinear functions h in these schemes lead to pseudorandom sequences such that their length are not maximal possible. We call such functions *unsuitable* and count the exact number of them for an arbitrary n.

II. Preliminaries

A *Boolean function in n variables* is a function of the form $f : \mathbb{F}_2^n \to \mathbb{F}_2$, where $\mathbb{F}_2^n = \{0, 1\}^n$ is a set of all binary vectors of length n.A *linear feedback shift register (LFSR)* consists of two parts: a binary vector

$$x = (x_{n-1}, \dots x_0)$$

of length n and a linear state function

$$f : (x_{n-1}, \dots, x_0) \to \{0, 1\},$$

where f is a Boolean function in n variables. A *state* of the register is a filling of vector x. During the encryption the register is changing its state using the feedback function. *Gamma* is a pseudorandom sequence generated by LFSR. A *period* is a length of repeating part of gamma.

Also, LFSR can be specified using feedback polynomials. It is a polynomial of degree n defining bits to be summed. If

$$f(x_{n-1}, \dots, x_0) = a_0 x_{n-1} \oplus a_1 x_{n-2} \oplus \dots \oplus a_{n-1} x_0,$$

than the corresponding feedback polynomial is defined as

$$p(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + 1.$$

If $p(z)$ is a primitive polynomial, i.e. the primitive element of the field $GF(2^n)$ is its root, then the period of a pseudorandom sequence generated by LFSR is maximal, i.e. is equal to $2^n - 1$. Thererfore, linear feedback shift registers are usually considers with primitive polinomials.

III. The analysis of gamma for linear feedback shift register generators

A. Filter generators

The filter generator consists of a single shift register of length n with a linear feedback and uses a primitive polynomial to change states. A Boolean function $h(x_{n-1}, \dots, x_0)$ generates a pseudorandom sequence gamma.

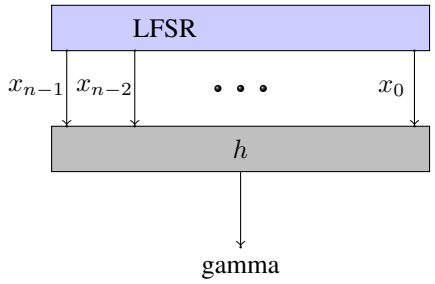


Fig. 1. Filter generator.

The work of the filter generator is shown in fig. 1, [8]. Let be $\gamma = (y_1 y_2 \dots y_{2^n-1})$, where $h(x_{n-1}, \dots, x_0) = y_1, h(x_{n-2}, \dots, x_0, f(x_{n-1}, \dots, x_0)) = y_2$, etc.. Since the state $x_0 = (0, \dots, 0)$ is not used, the number of all states equals $2^n - 1$. A Boolean function can generate gamma with the period from 1 to $2^n - 1$. In this paper we study how does the choice of the Boolean function h affect on periodic properties of generated gamma. Namely, we would like do determine all Boolean function h in n variables that lead to gammas with non-maximum period. Let us call such functions *unsuitable*. Note that the number of Boolean function does not depend on the linear state function. We give examples of suitable and unsuitable functions. Let $n = 4$ be a length of shift register, $p(z) = z^4 + z^3 + 1$ be primitive polynomial, $h_1 = x_0x_1x_2x_3 \oplus x_0x_2x_3 \oplus x_1x_3 \oplus x_1x_2 \oplus x_0x_3 \oplus x_3 \oplus x_2 \oplus x_1 \oplus 1$ and $h_2 = x_2x_3 \oplus x_1x_2 \oplus x_0x_3 \oplus x_3 \oplus x_1 \oplus 1$ be Boolean functions.

TABLE I
EXAMPLES OF SUITABLE AND UNSUITABLE FUNCTIONS

conditions	$h_1(x_3, x_2, x_1, x_0)$	$h_2(x_3, x_2, x_1, x_0)$
0001	1	1
0010	0	0
0100	0	1
1001	1	1
0011	0	0
0110	0	1
1101	1	0
1010	0	1
0101	0	1
1011	1	0
0111	0	1
1111	0	0
1110	1	1
1100	0	1
1000	0	0

As we see, h_1 generates gamma with period equals 3 and h_2 generates gamma with period equals 15.

Theorem 1. *Let n be an integer and*

$$2^n - 1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s}$$

where p_i are distinct prime numbers, $\alpha_i > 0$, s is an integer. Then the number of unsuitable Boolean functions in n variables for filter generators is equal to

$$2 \sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$.

Proof. Since $2^n - 1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s}$ the sequence can be divided into blocks of length $p_1^{i_1} p_2^{i_2} \dots p_s^{i_s}$, where $i_j = 0, \dots, \alpha_j$ and $\sum_{j=1}^s i_j < \sum_{j=1}^s \alpha_j$. Let us denote such blocks as $B_{i_1, i_2, \dots, i_s} = (y_1 y_2 \dots y_{p_1^{i_1} p_2^{i_2} \dots p_s^{i_s}})$. Let the degree of a block be denoted as $degree(B_{i_1, i_2, \dots, i_s}) = \sum_{j=1}^s i_j$. For blocks $B_{i_1, i_2, \dots, i_s}$ and $B_{j_1, j_2, \dots, j_s}$, where $gcd(p_1^{i_1} p_2^{i_2} \dots p_s^{i_s}, p_1^{j_1} p_2^{j_2} \dots p_s^{j_s}) = p_1^{k_1} p_2^{k_2} \dots p_s^{k_s}$, let us call block $B_{k_1, k_2, \dots, k_s}$ as a common block. The number of unsuitable sequences can be calculated using all possible blocks $B_{i_1, i_2, \dots, i_s}$ of degree less than $\sum_{j=1}^s \alpha_j$. Consider all such blocks $B_{i_1, i_2, \dots, i_s}$, where $\sum_{j=1}^s i_j = \sum_{j=1}^s (\alpha_j) - 1$. These blocks include all blocks of degree less than $\sum_{j=1}^s \alpha_j$. Let us count the number of blocks that are common for blocks of degree equal to $\sum_{j=1}^s (\alpha_j) - 1$. Thus, the number of unsuitable sequences can be calculated as

$$\sum_{i=1}^{s} (2^{p_1^{\alpha_1} \dots p_i^{\alpha_i - 1} \dots p_s^{\alpha_s}}) - A$$

where A is the number of blocks which are common for blocks of degree equal to $(\sum_{j=1}^s \alpha_j) - 1$.

For any two blocks $B_{\alpha_1, \dots, (\alpha_a - 1), \dots, \alpha_b, \dots, \alpha_s}$ and $B_{\alpha_1, \dots, \alpha_a, \dots, (\alpha_b - 1), \dots, \alpha_s}$ accordingly, there will be a common block $B_{\alpha_1, \dots, (\alpha_a - 1), \dots, (\alpha_b - 1), \dots, \alpha_s}$. Thus, the number of unsuitable sequences can be calculated as

$$\sum_{i=1}^{s} (2^{p_1^{\alpha_1} \dots p_i^{\alpha_i - 1} \dots p_s^{\alpha_s}}) - \sum_{1 \leq i < j \leq s} 2^{p_1^{\alpha_1} \dots p_i^{(\alpha_i - 1)} \dots p_j^{(\alpha_j - 1)} \dots p_s^{\alpha_s}} + B$$

where B is the number of blocks which are common for blocks of degree equal to $(\sum_{j=1}^s \alpha_j) - 2$. So, we continue in this way until we reach a common block between all blocks, namely $B_{(\alpha_1 - 1), (\alpha_2 - 1), \dots, (\alpha_s - 1)}$, and its degree is equal to $(\sum_{j=1}^s \alpha_j) - s$. Therefore, the number of unsuitable sequences is

$$2^{p_1^{(\alpha_1 - 1)} p_2^{\alpha_2} \dots p_s^{\alpha_s}} + 2^{p_1^{\alpha_1} p_2^{(\alpha_2 - 1)} \dots p_s^{\alpha_s}} + \dots + 2^{p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{(\alpha_s - 1)}} -$$
$$- 2^{p_1^{(\alpha_1 - 1)} p_2^{(\alpha_2 - 1)} p_3^{\alpha_3} \dots p_s^{\alpha_s}} - 2^{p_1^{(\alpha_1 - 1)} p_2^{\alpha_2} p_3^{(\alpha_3 - 1)} \dots p_s^{\alpha_s}} -$$
$$\dots - 2^{p_1^{\alpha_1} p_2^{\alpha_2} \dots p_{s-1}^{(\alpha_{s-1} - 1)} p_s^{(\alpha_s - 1)}} + \dots$$
$$+ (-1)^{s-1} 2^{p_1^{(\alpha_1 - 1)} p_2^{(\alpha_2 - 1)} \dots p_s^{(\alpha_s - 1)}}.$$

We can write all states of our register one by one: from one state we get the second one as the next state. Consider the vector of values of a Boolean function h that generates our gamma. Since in our set of states there is no zero state (it generates the cycle of length 1), our function h can take any value (0 or 1) on the zero vector. That is why there are exactly two Boolean functions that generate the same sequence.

So, the number of unsuitable functions is

$$2(2^{p_1^{(\alpha_1-1)}p_2^{\alpha_2}\dots p_s^{\alpha_s}} + 2^{p_1^{\alpha_1}p_2^{(\alpha_2-1)}\dots p_s^{\alpha_s}} + \dots + 2^{p_1^{\alpha_1}p_2^{\alpha_2}\dots p_s^{(\alpha_s-1)}} - 2^{p_1^{(\alpha_1-1)}p_2^{(\alpha_2-1)}p_3^{\alpha_3}\dots p_s^{\alpha_s}} - 2^{p_1^{(\alpha_1-1)}p_2^{\alpha_2}p_3^{(\alpha_3-1)}\dots p_s^{\alpha_s}} - \dots - 2^{p_1^{\alpha_1}p_2^{\alpha_2}\dots p_{s-1}^{(\alpha_{s-1}-1)}p_s^{(\alpha_s-1)}} + \dots + (-1)^{s-1}2^{p_1^{(\alpha_1-1)}p_2^{(\alpha_2-1)}\dots p_s^{(\alpha_s-1)}}). \quad \square$$

B. Combiner generators

Combiner generators use several linear feedback shift registers and each register has its own length n_i and use its own primitive polynomial for changing states. A Boolean function $h(X_{m-1}, \dots, X_0)$ where X_i is a register bit string i which generates pseudorandom sequence gamma. The work of the

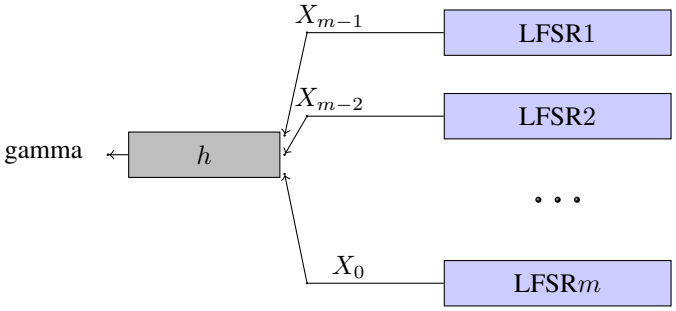


Fig. 2. Combiner generator.

combiner generator is shown in fig.2, [8]. Since we do not use the zero state the total number of states does not exceed $(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$. In this case, the maximum is reached at $gcd(n_i, n_j) = 1$ where $i, j = 1, \dots, m$, $i \neq j$. Then the Boolean function can generate a gamma with the period from 1 to $(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$. Boolean functions h in n variables leading to gammas generated non-maximum period are called unsuitable. Show that $gcd(2^{n_i}-1, 2^{n_j}-1) = 1$ where $i, j = 1, \dots, m$, $i \neq j$.

$$(2^{n_j}-1, 2^{n_i}-1) = (2^{n_j}-2^{n_i}, 2^{n_i}-1) =$$

$$= (2^{n_j-n_i}-1, 2^{n_i}-1) = 2^{(n_i,n_j)}-1 = 2^1-1 = 1.$$

For better understanding [2]. It means that each factor $(2^{n_i}-1)$ can be divided into

$$p_{k_1}^{\alpha_{k_1}} p_{k_2}^{\alpha_{k_2}} \dots p_{k_s}^{\alpha_{k_s}}.$$

Theorem 2. *Let n be an integer,*

$$\sum_{i=1}^{m} n_i = n,$$

where $i, j = 1, \dots, m$, $i \neq j$. And

$$(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1) = p_1^{\alpha_1}p_2^{\alpha_2}\dots p_s^{\alpha_s},$$

where p_i are different prime numbers, $\alpha_i > 0$, s is an integer. Then the number of unsuitable Boolean functions in n variables for the combiner generators is equal to

$$2^{2^{n_1+n_2+\dots+n_m}-(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)} \cdot \sum_{\beta\in\mathbb{F}_2^s,\beta\neq 0} ((-1)^{\beta_1+\dots+\beta_s+1}2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$.

Proof. Number of sequences for the combiner generators is equal to $\sum_{\beta\in\mathbb{F}_2^s,\beta\neq 0}((-1)^{\beta_1+\dots+\beta_s+1}2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}})$. Proof of this is similar to proof of number of sequences for the filter generators in theorem 1. As we use only $(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$ states and a total number of states is equal to $2^{n_1}\cdot 2^{n_2}\dots 2^{n_m} = 2^{n_1+n_2+\dots+n_m}$, then we have $2^{n_1+n_2+\dots+n_m}-(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)$ states, where our function can be equal 0 or 1. Therefore, for one this state we have two functions. In this way, the number of unsuitable Boolean functions in n variables for the combiner generators is equal to

$$2^{2^{n_1+n_2+\dots+n_m}-(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)} \cdot \sum_{\beta\in\mathbb{F}_2^s,\beta\neq 0} ((-1)^{\beta_1+\dots+\beta_s+1}2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$. $\square$

REFERENCES

[1] N. Tokareva, "Symmetric cryptography: a short course," Novosibirsk State University,pp.123–125, 2012.

[2] I. Vinogradov, "Basics of theory numbers," Nauka,pp.7–12, 1972.

[3] V. Roman'kov, "Introduction in cryptography," Lecture course, Forum, Moscow, pp.139–141, 2012.

[4] M. Gluhov, V. Elizarov, A. Nechaev, "Algebra ,",vol. 2, Gelios ARV,pp.327–333 ,2003.

[5] J. Golić, "On the security of nonlinear filter generators. In: Fast Software Encryption," Lecture notes in computer science, vol 1039. Springer, Heidelberg, pp.173–188, 1996.

[6] N. Courtois, W. Meier "Algebraic attacks on stream ciphers with linear feedback. In: Advances in cryptology - EUROCRYPT 2003," Lecture notes in computer science, vol 2656. Springer, Heidelberg, 2003

[7] E. Key, "An analysis of the structure and complexity of nonlinear binary sequence generators," IEEE Trans Inform Theory 22,pp.732–736, 1976.

[8] Carlet C. Boolean functions for cryptography and error-correcting codes // Boolean Models and Methods in Mathematics, Computer Science, and Engineering / Eds. P. Hammer, Y. Crama. Cambridge Univ. Press, 2010. Chapter 8. P. 257–397. URL: www.math.univ-paris13.fr/carlet/

On secondary and cyclic constructions of quadratic APN functions

Konstantin Kalgin*
Sobolev Institute of Mathematics
Novosibirsk State University
Novosibirsk, Russia
kalginkv@gmail.com

Valeriya Idrisova
Sobolev Institute of Mathematics
Novosibirsk, Russia
vvitkup@yandex.ru

Abstract

Almost perfect nonlinear functions possess the optimal resistance to the differential cryptanalysis and are widely studied. Most known construction of APN functions are obtained as functions over finite fields $\mathbb{F}_{2^n}$ and very little is known about combinatorial constructions in $\mathbb{F}_2^n$. In this work we proposed two approaches for constructing quadratic APN functions in $\mathbb{F}_2^n$. The first approach exploits a secondary construction idea, it considers how to obtain quadratic APN function in $n+1$ variables from a given quadratic APN function in n variables using special restrictions on new terms. The second approach is searching quadratic APN functions that have matrix form partially filled with standard basis vectors in a cyclic manner. Also, we conjectured that a quadratic part of an arbitrary APN function has a low differential uniformity. This conjecture allowed us to introduce a new subclass of APN functions, so-called stacked APN functions. We found cubic examples of such functions for dimensions up to 6.

1 Introduction

Let us recall some definitions. Let $\mathbb{F}_2^n$ be the n-dimensional vector space over $\mathbb{F}_2$. A function F from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$, where n and m are integers is called a *vectorial Boolean function*. If $m = 1$ such a function is called *Boolean*. Every vectorial Boolean function F can be represented as a set of m *coordinate functions* $F = (f_1, \ldots, f_m)$, where f_i is a Boolean function in n variables. Any vectorial function F can be represented uniquely in its *algebraic normal form (ANF)*:

$$F(x) = \sum_{I \in \mathcal{P}(N)} a_I (\prod_{i \in I} x_i),$$

*The work was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by Russian Foundation for Basic Research (projects no. 18-07-01394 and 20-31-70043) and Laboratory of Cryptography JetBrains Research.

where $\mathcal{P}(N)$ is a power set of $N = \{1, \ldots, n\}$ and $a_I \in \mathbb{F}_2^m$. The *algebraic degree* of a given function F is the degree of its ANF: deg (F) =max$\{|I| : a_I \neq 0, I \in \mathcal{P}(N)\}$. If algebraic degree of a function F is not more than 1 then F is called *affine*. If for an affine function F it holds $F(\mathbf{0}) = \mathbf{0}$ then F is called *linear*. If algebraic degree of a function F is equal to 2 then F is called *quadratic*.

Two vectorial functions F and G are *extended affinely equivalent (EA-equivalent)* if $F = A_1 \circ G \circ A_2 + A$ where A_1, A_2 are affine permutations on $\mathbb{F}_2^n$ and A is an affine function. Two functions F and G are called *Carlet-Charpin-Zinoviev [6] equivalent (CCZ-equivalent)* if their graphs $\{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \mid y = F(x)\}$ and $\{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \mid y = G(x)\}$ are affinely equivalent, that is, if there exists an affine automorphism $A = (A_1, A_2)$ of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ such that $y = F(x) \Leftrightarrow A_2(x, y) = G(A_1(x, y))$.

Let F be a vectorial Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$. For vectors $a, b \in \mathbb{F}_2^n$, where $a \neq 0$, consider the value

$$\delta(a, b) = \big|\{\ x \in \mathbb{F}_2^n \mid F(x + a) + F(x) = b\}\big|.$$

Denote by Δ_F the following value:

$$\Delta_F = \max_{a \neq \mathbf{0},\ b \in \mathbb{F}_2^n} \delta(a, b).$$

Then F is called *differentially Δ_F-uniform* function. The smaller the parameter Δ_F is the better the resistance of a cipher containing F as an S-box to differential cryptanalysis. For the vectorial functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ the minimal possible value of Δ_F is equal to 2. In this case the function F is called *almost perfect nonlinear (APN)*. This notion was introduced by K. Nyberg in [8].

APN fuctions draw attention of many researchers, but there is still a significant list [5] of important open questions, such as lower and upper bounds on the number of APN functions, an upper bound on algebraic degree of an APN function [4], the existence of bijective APN functions in even dimensions, etc. We are especially interested in two open problems that are devoted to constructing APN functions. The first one is to find secondary constructions of APN functions, in particular, it was stated as Problem 3.8 in [5]. The second problem is to find new constructions of APN functions in vectorspace $\mathbb{F}_2^n$, since almost all the known constructions of this class are found only as polynomials over the finite fields, and to the best of our knowledge, the only approach to such combinatorial constructions was proposed in [7].

In this work we propose two approaches for generating quadratic APN functions in $\mathbb{F}_2^n$. The first approach considers the algebraic normal form of a given quadratic APN function G in n variables and extends it into an ANF of a quadratic function F in $n + 1$ variables, using special restrictions on coefficients of new terms. In the second method we consider special matrices that are partially filled with vectors of standard basis and search for corresponding APN functions using the same idea of restrictions. Generally, quadratic APN functions are not suitable as secure S-boxes due to the low algebraic degree, but obtaining new quadratic representatives can lead us to another useful functions. This is very important for even $n \geqslant 8$, since new APN permutations CCZ-equivalent to quadratic functions can be found for these dimensions [3].

In the last part of the work we conjectured that a quadratic part of an arbitrary APN function has a low differential uniformity. We introduced the new notion of stacked APN function and for dimensions up to 6 found such functions using quadratic APN functions obtained with approaches mentioned above.

2 On secondary construction of quadratic APN functions

Since EA-equivalence preserves APNness, it is always possible to omit linear and constant terms in the algebraic normal form of a given APN function. Further we will consider quadratic vectorial Boolean functions that have only quadratic terms in their ANF. The following theorem gives a necessary condition on the ANF of a given APN function.

Theorem 1. *[1] Let $F = (f_1, \dots, f_n)$ be an APN function in n variables. Then every quadratic term $x_i x_j$, where $i \neq j$, appears at least in one coordinate function of F.*

This property motivated us to suggest the following construction of quadratic APN functions. Let $G = (g_1, \dots, g_n)$ be a quadratic APN-function in n variables. Consider vectorial function $F = (f_1, \dots, f_n, f_{n+1})$ in $n+1$ variables such that:

$$\begin{aligned} f_1 &= g_1 + \sum_{i=1}^{n} \alpha_{1,i} x_i x_{n+1}; \\ &\dots \\ f_n &= g_n + \sum_{i=1}^{n} \alpha_{n,i} x_i x_{n+1}; \\ f_{n+1} &= g_{n+1} + \sum_{i=1}^{n} \alpha_{n+1,i} x_i x_{n+1}, \end{aligned} \tag{1}$$

where $\alpha_{1,i} \dots, \alpha_{n+1,i} \in \mathbb{F}_2$ for $i = 1, \dots, n$ and $g_{n+1} = \sum_{1 \leqslant j < k \leqslant n} \beta_{j,k} x_j x_k$ for some fixed $\beta_{j,k} \in \mathbb{F}_2$. Note that if $\alpha_{1,i}, \dots, \alpha_{n,i}$ are such that each term $x_i x_{n+1}$ appears at least in one of the coordinate functions $f_1, \dots, f_n$, then the necessary condition of Theorem 1 is held for the constructed function F. Since the exhaustive search for the given APN function becomes complicated starting from $n = 6$, there is a need to find necessary and sufficient conditions on new coefficients of F.

Let us denote the lexicographically ordered elements of $\mathbb{F}_2^n$ as $x^0, \dots, x^{2^n-1}$. Since all the values $G(x^0), \dots, G(x^{2^n-1})$ of function G are known, we can represent values of the constructed function F only through unknown coefficients $\alpha_{i,k}$ and some constant terms. Since F is an APN function, for a nonzero a all sums $F(x) + F(x+a)$ and $F(y) + F(y+a)$, where $x \neq y$ and $x \neq y + a$, should be pairwise different. This fact applies special restrictions on coefficients $\alpha_{i,k}$. For the convenient representation of these restrictions further we consider the following matrix approach that was proposed by Beth and Ding in [1].

Each quadratic vectorial function G in n variables can be considered as a symmetric matrix $\mathcal{G} = (g_{ij})$, where each element $g_{ij} \in \mathbb{F}_2^n$ is a vector of coefficients corresponding to term $x_i x_j$ in the algebraic normal form of G and all diagonal elements g_{ii} are null.

Example 2. For $n = 3$ let us consider function $G = (g_1, g_2, g_3) = (x_1x_2, x_2x_3, x_1x_3)$ $= \begin{bmatrix}1\\0\\0\end{bmatrix} \cdot x_1x_2 + \begin{bmatrix}0\\0\\1\end{bmatrix} \cdot x_1x_3 + \begin{bmatrix}0\\1\\0\end{bmatrix} \cdot x_2x_3.$

Then the corresponding matrix $\mathcal{G}$ is the following:

$$\mathcal{G} = \begin{bmatrix}(000) & (100) & (001)\\(100) & (000) & (010)\\(001) & (010) & (000)\end{bmatrix}$$

It is necessary to mention that these matrices also were used in [10] and [9] to construct and classify a lot of new quadratic APN functions over finite fields. Using these matrices the APN property can be formulated in the following way:

Proposition 3. *Let $\mathcal{G}$ be the matrix that corresponds to quadratic vectorial function G. Then function G is APN if and only if $x \cdot (\mathcal{G} \cdot a) \neq 0$ for all $x \neq a$, where $a, x \in \mathbb{F}_2^n$ and $a \neq 0$.*

In terms of matrices the construction from (1) can be considered as an extension of a given $\mathcal{G}$ with an extra bit that represents g_{n+1} in every element and an extra pair of row and column that represents a set of new terms x_ix_{n+1}.

Example 4. For the considered APN function $G = (g_1, g_2, g_3) = (x_1x_2, x_2x_3, x_1x_3)$ we choose null g_{n+1} and construct APN function $F = (f_1, f_2, f_3, f_4)$ in 4 variables, where:
$f_1 = g_1$;
$f_2 = g_2 + x_3x_4$;
$f_3 = g_3 + x_2x_4 + x_3x_4$;
$f_4 = x_1x_4 + x_3x_4$.
Then the corresponding matrix $\mathcal{F}$ is the following:

$$\mathcal{F} = \begin{bmatrix}(0000) & (1000) & (0010) & (0001)\\(1000) & (0000) & (0100) & (0010)\\(0010) & (0100) & (0000) & (0111)\\(0001) & (0010) & (0111) & (0000)\end{bmatrix}$$

Consider a quadratic APN function G and the corresponding $n \times n$ matrix $\mathcal{G}$. Denote the vector of nonzero coefficients as $\alpha = (\alpha_1, \ldots, \alpha_n)$. Let us fix g_{n+1} and construct $(n+1) \times (n+1)$ matrix $\mathcal{F}$ by adding $(\alpha_1, \ldots, \alpha_n, 0)$ as the last column and the last row and adding new bit to every element according to the choice of g_{n+1}. Let us denote as $\mathcal{G}'$ the submatrix (f_{ij}) of $\mathcal{F}$, such that $i, j < n+1$. Let $\langle X \rangle$ denote the linear span of X and F be the quadratic vectorial that is corresponded with the constructed matrix $\mathcal{F}$

Theorem 5. *A function F is APN if and only if $\alpha \cdot a'$ does not belong to $\langle \mathcal{G}' \cdot a' \rangle$ for all $a' \in \mathbb{F}_2^n$, $a' \neq \boldsymbol{0}$.*

Let us note that Theorem 5 shows how to obtain restrictions on new coefficients in the convenient form.

For the given $k \in \mathbb{N}$ let us consider the following sets:

$S_{i,k} = \{\alpha_i + v \mid v \in \langle \mathcal{G}' \cdot (e_i + e_k)\rangle\}$;

$S_{i,j,k} = \{\alpha_i + \alpha_j + v \mid v \in \langle \mathcal{G}' \cdot (e_i + e_j + e_k)\rangle\}$;

...

$S_{1,2,\ldots,k-1,k} = \{\alpha_1 + \alpha_2 + \ldots + \alpha_{k-1} + v \mid v \in \langle \mathcal{G}' \cdot (e_1 + e_2 + \ldots + e_{k-1} + e_k)\rangle\}$,

where $e_1, \ldots, e_n$ is the standard basis in $\mathbb{F}_2^n$. Let us call a vector $\alpha = (\alpha_1, \ldots, \alpha_n)$, where $\alpha_i \in \mathbb{F}_2^{n+1}$, *admissible* for matrix $\mathcal{G}'$ if it satisfies the condition in Theorem 5. We call a sequence $(\alpha_1^*, \ldots, \alpha_k^*)$, where $\alpha_i^* \in \mathbb{F}_2^{n+1}$, to be *$k$-admissible* for some $k \leqslant n$, if vector $\alpha^* = (\alpha_1^*, \ldots, \alpha_k^*, \mathbf{0}, \ldots, \mathbf{0})$ of length n is admissible for all nonzero $a' = (a'_1, \ldots, a'_n) \in \mathbb{F}_2^n$ such that $a'_{k+1} = 0, \ldots, a'_n = 0$. An n-admissible sequence can be considered as an admissible vector of length n. Consider an APN function G in n variables and a fixed g_{n+1}.

Proposition 6. *The number of quadratic APN functions that can be obtained from function G using the construction from (1) is equal to the number of admissible vectors $\alpha = (\alpha_1, \ldots, \alpha_n)$ for matrix $\mathcal{G}'$.*

It can be seen that there are $2^{n+1} - \mid \langle \mathcal{G}' \cdot (e_1)\rangle \mid$ vectors α_1 such that (α_1) is 1-admissible. The following proposition shows how to obtain the number of admissible vectors:

Proposition 7. *Let $(\alpha_1, \alpha_2, \ldots, \alpha_{k-1})$ be the $(k-1)$-admissible sequence for some $k < n+1$. Then there exist*

$$2^{n+1} - \mid \langle \mathcal{G}' \cdot (e_k)\rangle \cup \{\bigcup_{i=1}^{k-1} S_{i,k}\} \cup \{\bigcup_{1 \leqslant i < j < k,} S_{i,j,k}\} \cup \ldots \cup S_{1,2,\ldots,k-1,k} \mid$$

vectors α_k such that sequence $(\alpha_1, \alpha_2, \ldots, \alpha_{k-1}, \alpha_k)$ is k-admissible.

Also, our method can be extended to the case when G is not an APN function, but ANF of G and g_{n+1} together contain all possible quadratic terms. The following proposition describes the necessary condition on the choice of such functions.

Proposition 8. *Let G be a vectorial function in n variables and F be an APN function in $n+1$ variables that it is obtained from G using construction (1). Then $\Delta_G \leqslant 4$.*

For example, for differential 4-uniform function $G = (g_1, g_2, g_3, g_4, g_5)$, where:

$g_1 = x_1x_2 + x_3x_5 + x_4x_5$;

$g_2 = x_1x_3 + x_4x_5$;

$g_3 = x_2x_3 + x_1x_4 + x_3x_5 + x_4x_5$;

$g_4 = x_2x_4 + x_1x_5 + x_4x_5$;

$g_5 = x_3x_4 + x_2x_5 + x_4x_5$.

and g_6 contains all the terms $x_i x_j$, where $i < j \leqslant n$, we obtained 13 CCZ classes among constructed functions. Let us recall that there exist only 13 CCZ classes of quadratic APN functions in dimension 6.

It can be seen that every quadratic APN function can be obtained using construction from(1). It is worth mentioning that when $n = 3, 4$ and 5 for APN functions that are CCZ classes representatives we obtained all the possible classes of quadratic APN functions for $4, 5$ and 6 variables from the classification [2] and large variety of classes for constructing from 6 to 7 variables.

Note that for the given APN function G in n variables we have $2^{\frac{(n^2-n)}{2}}$ possibilities to choose g_{n+1}. It is interesting that the choice of g_{n+1} affects the capability to obtain APN function F in $n+1$ variables, the number of such constructed functions and the variety of different CCZ-classes among constructed classes. For example, when $n = 5$ and g_{n+1} is null both quadratic CCZ-representatives give us the only one CCZ-class for 6 variables. At the same time for another choices of g_{n+1} these functions give 13 CCZ-classes of quadratic APN functions in 6 variables. Unfortunately, for $n \geqslant 7$ it becomes computationally harder to choose the proper initial function and g_{n+1} and to obtain a large amount of generated functions. It seems that construction from (1) is not so efficient on large dimensions.

3 On cyclic construction of quadratic APN functions

Let us introduce another approach for constructing quadratic APN functions using matrix representation from previous section. Let $e_1, \ldots, e_n$ be the standard basis in $\mathbb{F}_2^n$. For the given n consider the following matrix with elements from $\mathbb{F}_2^n$:

$$\mathcal{T} = \begin{bmatrix} 0 & e_1 & e_2 & e_3 & \ldots & e_{n-2} & e_{n-1} \\ e_1 & 0 & e_3 & e_4 & \ldots & e_{n-1} & e_n \\ e_2 & e_3 & 0 & e_5 & \ldots & e_n & t_{3,n} \\ e_3 & e_4 & e_5 & 0 & \ldots & t_{4,n-1} & t_{4,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ e_{n-2} & e_{n-1} & e_n & t_{n-1,4} & \ldots & 0 & t_{n-1,n} \\ e_{n-1} & e_n & t_{n,3} & t_{n,4} & \ldots & t_{n,n-1} & 0 \end{bmatrix},$$

where $t_{i,j} = t_{j,i}$ and $t_{i,j}$ denote some unknown elements in $\mathbb{F}_2^n$.

Our aim is to find values of missed matrix elements such that matrix $\mathcal{T}$ represents APN function. We can apply the approach with restrictions from the previous section. Without loss of generality let us consider the first unknown element of matrix $\mathcal{T}$ that is $t_{3,n}$. According to Theorem 5 the last column of $\mathcal{T}$ should satisfy $(e_{n-1}, e_n, t_{3,n}, \ldots, 0) \cdot a' \notin \langle \mathcal{T}' \cdot a' \rangle$, where $a' \in \mathbb{F}_2^{n-1}$, $a' \neq 0$ and $\mathcal{T}' = \mathcal{T} \setminus (e_{n-1}, e_n, t_{3,n}, \ldots, 0)$. If we consider all $a' = a'_1, \ldots, a'_{n-1}$ such that $a'_3 = 1$ and $a'_i = 0$, if $i > 3$, we obtain restrictions on the value of $t_{3,n}$ that are independent from any other unknown element of $\mathcal{T}$. Repeating this procedure step by step for every new element after fixing values of previous variables $t_{i,j}$ allows us to obtain all possible fillings for the given matrix $\mathcal{T}$.

For $n = 3, 4$ and 5 this construction covered all quadratic CCZ-classes. For $n = 6$ it covered 11 out of 13 classes. Unfortunately, for larger dimensions the number of generated

functions dropped dramatically and the construction covers only 7 classes for $n = 7$ and only one class for $n = 8$. As a consequence, we consider the following generalization of this construction.

Let $\mathcal{T}$ be the same matrix that contains k unknown elements. Consider the diagonal that contains all elements e_n in $\mathcal{T}$. It is easy to see that we can remove any element e_n from this diagonal and apply the above procedure to the new matrix with $k+1$ unknown elements. Moreover, we can remove any number of elements e_n from this diagonal and the more elements are deleted the more APN functions can be constructed using this matrix.

For $n = 6$ when we removed one element e_n from the diagonal in $\mathcal{T}$ the new matrix had already covered all 13 CCZ classes of quadratic APN functions. For $n = 7$ and the matrix that has no elements e_n on the diagonal we generated 2341888 quadratic APN functions, the variety of CCZ classes is still being checking at the moment as well as generating quadratic APN functions for $n = 8$ is in progress. It is interesting to note that such construction generates, in some sense, lexicographically smallest quadratic functions.

4 The differential uniformity of APN functions quadratic parts and stacked APN functions

Let F be a vectorial Boolean function of algebraic degree d. Then it can be represented as sum $F = F^{(c)} + F^{(1)} + F^{(2)} + \ldots + F^{(d)}$, where each function $F^{(j)}$ contains only monomials of algebraic degree j and $F^{(c)}$ is a constant term. We observed that if F is an APN function then its quadratic part $F^{(2)}$ has a low differential uniformity.

Conjecture 9. Let F be an APN function in n variables, where $4 \leqslant n \leqslant 7$. Then $\Delta_{F^{(2)}} \leqslant 4$.

The conjecture is true for $n = 4$. When $n = 8, 9$ there were found APN functions F (e.g. Kasami power functions for $n = 8$ and Inverse function for $n = 9$) such that $\Delta_{F^{(2)}} = 8$. Nevertheless, for these large dimensions the differential uniformity of quadratic parts is still quite low. Further we consider only functions without affine terms. Let F be an APN function in n variables, where $F = F^{(2)} + F^{(3)} + \ldots + F^{(d)}$.

Proposition 10. *If* $H = F + F^{(2)} = (0, \ldots, 0, h_j, 0, \ldots, 0)$ *for some* $1 \leqslant j \leqslant n$, *then* $\Delta_{F^{(2)}} \leqslant 4$.

For $n = 4, 6$ there exist cubic APN functions such that $H = F + F^{(2)} = (0, \ldots, 0, h_j, 0, \ldots, 0)$ for some $1 \leqslant j \leqslant n$. Let us note that these simple results allow us to use quadratic APN or differential 4-uniform functions to construct functions of higher degrees, particularly, cubic APN functions.

The observation on low differential uniformity of an APN function quadratic part motivated us to introduce a new subclass of APN functions.

Definition 11. Let $F = F^{(2)} + \ldots + F^{(d)}$ be an APN function of algebraic degree d. If all functions $F + F^{(d)}$, $F + F^{(d)} + F^{(d-1)}, \ldots, F + F^{(d)} + F^{(d-1)} + \ldots + F^3$ are APN functions then F is called a *stacked APN function.*

Let us describe one of the possible approaches to constructing stacked APN functions of degree 3. Let H be a cubic vectorial function in n variables with no affine or quadratic terms. Then $H = \sum_{i,j,k} a_{ijk} x_i x_j x_k$, where $1 \leqslant i < j < k \leqslant n$ and $a_{ijk} \in \mathbb{F}_2^n$. Let $a_{i_1 j_1 k_1}$ be an arbitrary nonzero coefficient in the ANF of H. Let us call H a *cubic shift* if for all $1 \leqslant i < j < k \leqslant n$ vector a_{ijk} is null or equal to $a_{i_1 j_1 k_1}$. For $n = 4, 5$ we implemented the search of cubic APN functions $F = F^{(2)} + F^{(3)}$ such that $F^{(3)}$ is some cubic part and $F^{(2)}$ is an APN quadratic function, that is constructed using the cyclic matrix $\mathcal{T}$ from the previous section. For $n = 6$ we implemented the similar search, but $F^{(3)}$ was a cubic shift since it is computationally hard to search through all the possible cubic parts. We have found a large amount of cubic stacked APN functions for $n = 4, 5, 6$. Some examples are provided below:

Table 1: Examples of stacked cubic APN functions (both F and $F^{(2)}$ are APN).

F	0	0	0	1	0	2	4	7	0	4	6	3	8	14	11	12
$F^{(2)}$	0	0	0	1	0	2	4	7	0	4	6	3	8	14	10	13
F	0	0	0	1	0	2	4	7	0	4	10	15	19	21	28	27
	0	8	16	25	11	1	29	22	15	3	17	28	31	17	6	9
$F^{(2)}$	0	0	0	1	0	2	4	7	0	4	10	15	19	21	29	26
	0	8	16	25	11	1	31	20	15	3	21	24	23	25	9	6
F	0	0	0	1	0	2	4	13	0	4	8	7	16	22	28	27
	0	8	16	19	9	3	29	22	45	33	53	56	52	58	40	45
	0	16	60	45	26	8	34	59	55	35	3	28	61	43	13	26
	5	29	41	58	22	12	62	37	31	3	59	38	28	2	60	41
$F^{(2)}$	0	0	0	1	0	2	4	7	0	4	8	13	16	22	28	27
	0	8	16	25	9	3	29	22	45	33	53	56	52	58	40	39
	0	16	60	45	26	8	34	49	55	35	3	22	61	43	13	26
	5	29	41	48	22	12	62	37	31	3	59	38	28	2	60	35

It is worth mentioning that we have found more than 70 000 cubic stacked APN functions when $n = 6$ and all these functions belong to the same CCZ-class that is the only known class that does not contain quadratic functions (the class number 13 in the list from [2]).

There have been left a few open questions:

1. How to choose properly the initial function G in secondary approach? It seems that for most APN functions in n variables it is possible to find corresponding APN functions in $n + 1$ variables for some g_{n+1}, but we have found one counterexample when $n = 6$.

2. The same question arised for the choice of g_{n+1}, since it greatly affects the outcome of the search.

3. It is possible to extend the above idea of restrictions on larger algebraic degree cases, in particular, on cubic case. How to find a convenient systematic way of obtaining these restrictions (as it was made with matrices in quadratic case)?

4. Conjecture 9 can be extended for larger dimensions, i.e. $\Delta_{F^{(2)}} \leqslant 8$ when $8 \leqslant n \leqslant N_1$ for some N_1. Can we estimate value N_1?

5. Is there exist stacked APN functions of degree more than 3?

Acknowledgements

We would like to cordially thank Natalia Tokareva for her valuable remarks and careful reading. The work was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by Russian Foundation for Basic Research (projects no. 18-07-01394 and 20-31-70043) and Laboratory of Cryptography JetBrains Research.

References

[1] T. Beth, C. Ding. On almost perfect nonlinear permutations. *Advances in Cryptology, EUROCRYPT'93, Lecture Notes in Computer Science*, vol. 765, pp. 65-76, 1993.

[2] M. Brinkmann, G. Leander. On the classification of APN functions up to dimension five. *Des. Codes Cryptogr.*, vol. 49, Issue 13, pp. 273-288, 2008.

[3] K. A. Browning, J. F. Dillon, M. T. McQuistan, A. J. Wolfe. An APN Permutation in Dimension Six. *Post-proceedings of the 9-th International Conference on Finite Fields and Their Applications Fq'09, Contemporary Math., AMS*, vol. 518, pp. 33-42, 2010.

[4] L. Budaghyan, C. Carlet, T. Helleseth, N. Li and B. Sun. On Upper Bounds for Algebraic Degrees of APN Functions. *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4399-4411, 2018.

[5] C. Carlet. Open Questions on Nonlinearity and on APN Functions. *Arithmetic of Finite Fields. WAIFI 2014. Lecture Notes in Computer Science*, vol. 9061, pp 83-107 (2015).

[6] C. Carlet, P. Charpin, V. Zinoviev. Codes, bent functions and permutations suitable for DES-like cryptosystems. *Des. Codes Cryptogr.*, vol. 15, pp. 125-156, 1998.

[7] A. A. Gorodilova. Characterization of almost perfect nonlinear functions in terms of subfunctions, *Diskr. Mat.*, vol. 27(3), pp. 3-16 (2015); *Discrete Math. Appl.*, vol. 26(4), pp. 193-202, 2016.

[8] K. Nyberg. Differentially uniform mappings for cryptography. *Advances in Cryptography, EUROCRYPT'93, Lecture Notes in Computer Science*, vol. 765, pp. 55-64, 1994.

[9] Y. Yu, N. S. Kaleyski, L. Budaghyan, Y. Li. Classification of quadratic APN functions with coefficients in GF(2) for dimensions up to 9. *IACR Cryptol. ePrint Arch.*: 1491, 2019.

[10] Y. Yu, M. Wang, Y. Li. A matrix approach for constructing quadratic APN functions. *Des. Codes Cryptogr.* 73, 587-600, 2014

On constructions and properties of self-dual generalized bent functions

Aleksandr Kutsenko*

Sobolev Institute of Mathematics, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia

alexandrkutsenko@bk.ru

Abstract

Bent functions of the form $\mathbb{F}_2^n \to \mathbb{Z}_q$, where $q \geqslant 2$ is a positive integer, are known as generalized bent (gbent) functions (K.-U Schmidt, 2006). There is a class of gbent functions for which it is possible to define a dual gbent function, gbent functions that possess this property are called regular. A regular gbent function is said to be self-dual if it coincides with its dual. In this paper we explore self-dual generalized bent functions. We give necessary and sufficient conditions for the self-duality of Maiorana–McFarland gbent functions, consider self-dual bent functions obtained by the direct sum of generalized Boolean functions. We provide a sufficient condition for a gbent function from Dillon's Partial Spreads to be self-dual. Two iterative constructions based on the generalization of iterative constructions of Boolean self-dual bent functions are presented. We prove that the set of sign functions of self-dual gbent functions in n variables has dimension 2^{n-1}. We find all self-dual gbent functions symmetric with respect to two variables and prove that self-dual gbent function can not be affine. Symmetries that preserve self-duality are also discussed.

1 Introduction

Boolean bent functions were introduced by [19], they have applications in cryptograohy and coding theory. In 2000, Wada [28] established a connection between bent functions and binary constant-amplitude codewords.

Having applications of functions from $\mathbb{F}_2^n$ to $\mathbb{Z}_4$ in code-division multiple access (CDMA) systems, K.-U Schmidt introduced in [20] the bentness of a generalized Boolean function and also studied quaternary generalized bent (gbent) functions (see also paper [21]). Note that generalized Boolean functions were also studied in a perspective of obtaining linear codes, see [15]. In recent years generalized bent functions obtained much attention. In

*The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

papers [13, 23] several constructions and properties of generalized bent functions were obtained. The question of the characterization of generalized bent functions was recently studied in [7, 14, 24].

Self-dual bent functions were explored by C. Carlet et al. in 2010 [3], main constructions and properties were given and the classification for small number of variables was provided. Next steps for the classification were made in [5], quadratic self-dual bent functions were characterized in [9]. Other constructions, metrical properties and groups of automorphisms of self-dual bent functions were studied in [10, 11, 12]. In 2018 L. Sok. et al. in paper [22] studied quaternary self-dual bent functions from the viewpoints of existence, construction, and symmetry. In current work we investigate constructions, symmetries and other properties of self-dual generalized bent functions $\mathbb{F}_2^n \to \mathbb{Z}_q$, when q is even.

A survey on different generalizations of bent functions can be found in [25].

2 Notation

Let $\mathbb{F}_2^n$ be a set of binary vectors of length n. For $x, y \in \mathbb{F}_2^n$ denote $\langle x, y\rangle = \bigoplus_{i=1}^{n} x_i y_i$, where the sign $\oplus$ denotes a sum modulo 2.

A *generalized Boolean function* f in n variables is any map from $\mathbb{F}_2^n$ to $\mathbb{Z}_q$, the integers modulo q. The set of generalized Boolean functions in n variables is denoted by $\mathcal{GF}_n^q$. Let $\omega = e^{2\pi i/q}$. A *sign* function of $f \in \mathcal{GF}_n^q$ is a complex valued function ω^f, we will also refer to it as to a complex vector $\left(\omega^{f_0}, \omega^{f_1}, ..., \omega^{f_{2^n-1}}\right)$ of length 2^n, where $(f_0, f_1, ..., f_{2^n-1})$ is a vector of values of the function f.

The *Hamming weight* $\mathrm{wt}_H(x)$ of the vector $x \in \mathbb{F}_2^n$ is the number of nonzero coordinates of x. The *Hamming distance* $\mathrm{dist}_H(f, g)$ between generalized Boolean functions f, g in n variables is the cardinality of the set $\{x \in \mathbb{F}_2^n | f(x) \neq g(x)\}$. The Lee weight of the element $x \in \mathbb{Z}_q$ is $\mathrm{wt}_L(x) = \min\{x, q - x\}$. The Lee distance $\mathrm{dist}_L(f, g)$ between $f, g \in \mathcal{GF}_n^q$ is

$$\mathrm{dist}_L(f, g) = \sum_{x \in \mathbb{F}_2^n} \mathrm{wt}_L\left(\delta(x)\right),$$

where $\delta \in \mathcal{GF}_n^q$ and $\delta(x) = f(x) + (q-1)g(x)$ for any $x \in \mathbb{F}_2^n$. For Boolean case $q = 2$ the Hamming distance coincides with the Lee distance.

The *(generalized) Walsh–Hadamard transform* of $f \in \mathcal{GF}_n^q$ is the complex valued function:

$$H_f(y) = \sum_{x \in \mathbb{F}_2^n} \omega^{f(x)} (-1)^{\langle x, y\rangle}.$$

A generalized Boolean function f in n variables is said to be *generalized bent* (gbent) if

$$|H_f(y)| = 2^{n/2},$$

for all $y \in \mathbb{F}_2^n$ [20]. If there exists such $\widetilde{f} \in \mathcal{GF}_n^q$ that $H_f(y) = \omega^{\widetilde{f}(y)} 2^{n/2}$ for any $y \in \mathbb{F}_2^n$, the gbent function f is said to be *regular* and $\widetilde{f}$ is called its *dual*. Note that $\widetilde{f}$ is generalized

bent as well. A regular gbent function f is said to be *self-dual* if $f = \widetilde{f}$, and *anti-self-dual* if $f = \widetilde{f} + \frac{q}{2}$. Consequently, it is the case only for even q. So throughout this paper we assume that q is a natural even number.

3 Constructions

3.1 Direct sum

Suppose $n = n_1 + n_2 + \cdots + n_r$ and $p_k \leqslant q$, where n_k, p_k are positive integers for $k = 1, 2, ..., r$. Let $f \in \mathcal{GF}_n^q$, consider gbent functions $f_k \in \mathcal{GF}_n^{p_k}$, $k = 1, 2, ..., r$. The function

$$f(x) = f_1\left(x^{(1)}\right) + f_2\left(x^{(2)}\right) + \cdots + f_r\left(x^{(r)}\right),$$

where $x^{(k)} \in \mathbb{F}_2^{n_k}$ and $x = \left(x^{(1)}, x^{(2)}, ..., x^{(r)}\right) \in \mathbb{F}_2^n$, is a *direct sum* of generalized Boolean functions f_k. Gbent functions obtained by a direct sum of generalized Boolean functions were studied in paper [8], it was proved that function f is gbent if and only if all f_k are gbent functions. Here we consider self-dual bent functions obtained by this construction.

Proposition 1. *Assume all numbers p_k are even and $f_k \in \mathcal{GF}_n^{p_k}$ are gbent functions such that $\widetilde{f_k} = f_k + c_k\,(p_k/2)$, where $c_k \in \mathbb{F}_2$, $k = 1, 2, ..., r$. If there is an even number of nonzero coefficients c_k, the function f is a self-dual gbent function in n variables.*

3.2 Maiorana–McFarland class

Bent functions in $2k$ variables which have a representation

$$f(x, y) = \langle x, \pi(y)\rangle \oplus g(y),$$

where $x, y \in \mathbb{F}_2^k$, $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and g is a Boolean function in k variables, form the well known *Maiorana–McFarland* class of bent functions. It is known [2] that a dual of a Maiorana–McFarland bent function $f(x, y)$ is equal to

$$\widetilde{f}(x, y) = \langle \pi^{-1}(x), y\rangle \oplus g\left(\pi^{-1}(x)\right).$$

A generalization of this construction for the case $q = 4$ was given by Schmidt in [20]. In [23] this construction was given for any even q, thus, forming the following construction

$$f(x, y) = \frac{q}{2}\langle x, \pi(y)\rangle + g(y),$$

where $x, y \in \mathbb{F}_2^k$, $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and g is a generalized Boolean function in k variables. Its dual is

$$\widetilde{f}(x, y) = \frac{q}{2}\langle \pi^{-1}(x), y\rangle + g\left(\pi^{-1}(x)\right).$$

In the article [3] necessary and sufficient conditions of (anti-)self-duality of Maiorana–McFarland bent functions, denoted by $\mathrm{SB}^+_{\mathcal{M}}(n)$ ($\mathrm{SB}^-_{\mathcal{M}}(n)$), were given. In [22] quaternary

self-dual Maiorana–McFarland bent functions were studied and necessary and sufficient conditions of self-duality were obtained for them.

In the current work we generalize these results for any even q. Denote the sets of (anti-)self-dual generalized Maiorana–McFarland bent functions by $\mathrm{SB}^{+}_{\mathcal{M}^q}(n)$ ($\mathrm{SB}^{-}_{\mathcal{M}^q}(n)$)

Theorem 2. *A generalized Maiorana–McFarland bent function*

$$f(x,y)=\frac{q}{2}\langle x,\pi(y)\rangle+g(y),\ x,y\in\mathbb{F}_2^{n/2},$$

is (anti-)self-dual bent if and only if for any $y\in\mathbb{F}_2^{n/2}$

$$\pi(y)=L\left(y\oplus b\right),\quad g(y)=\frac{q}{2}\langle b,y\rangle+d,$$

where $L\in\mathcal{O}_{n/2}, b\in\mathbb{F}_2^{n/2}$, $\mathrm{wt}\,(b)$ is even (odd), $d\in\mathbb{Z}_q$.

It follows that the number of such functions is a function of q and the cardinality of the orthogonal group.

Corollary 3. *It holds*

$$\left|\mathrm{SB}^{+}_{\mathcal{GM}^q}(n)\right|=\left|\mathrm{SB}^{-}_{\mathcal{GM}^q}(n)\right|=q\cdot 2^{n/2-1}\left|\mathcal{O}\left(n/2,\mathbb{F}_2\right)\right|.$$

3.3 Dillon functions type

In [13] an explicit representation of functions in a generalization of Dillon's $\mathcal{PS}_{ap}$ class to gbent functions with $q=2^k$ was presented. By comparing the function from $\mathcal{PS}_{ap}$ in a bivariate form with its dual (that was also given in [13]) we obtain the following result.

Theorem 4. *Assume G_j, $j=0,1,...,k-1$, be balanced Boolean functions in m variables with $G_j(0)=0$ and $\sum\limits_{t\in\mathbb{F}_{2^m}}\omega^{\sum\limits_{j=0}^{k-1}2^jG_j(t)}=0$. Then, if $G_j(u)=G_j(1/u)$ for any $u\in\mathbb{F}_{2^m}$ (with the convention $1/0=0$), then the function $f:\mathbb{F}_{2^m}\times\mathbb{F}_{2^m}\to\mathbb{Z}_{2^k}$ given by*

$$f(x,y)=\sum_{j=0}^{k-1}2^jG_j(x/y)$$

is self-dual gbent in $2m$ variables.

3.4 Iterative construction

Let f_0,f_1,f_2,f_3 be Boolean functions in n variables. Consider a Boolean function g in $n+2$ variables which is defined as

$$g(00,x)=f_0(x),\quad g(01,x)=f_1(x),\quad g(10,x)=f_2(x),\quad g(11,x)=f_3(x),\quad x\in\mathbb{F}_2^n.$$

It is known (Preneel et al., 1991; see also [1, 26]) that under condition that all f_0, f_1, f_2, f_3 are Boolean bent functions in n variables, the mentioned function g is a bent function in $n+2$ variables if and only if

$$\widetilde{f}_0 \oplus \widetilde{f}_1 \oplus \widetilde{f}_2 \oplus \widetilde{f}_3 = 1,$$

that gives the construction of a bent function in $n+2$ variables through the concatenation of vectors of values of four bent functions in n variables [16].

Following N. Tokareva [26], we will refer to Boolean bent functions obtained by this construction as *bent iterative functions* ($\mathcal{BI}$) the set of such bent functions in n variables is denoted by $\mathcal{BI}_n$. A construction of generalized bent functions in $n+2$ variables obtained by concatenation of four generalized Boolean functions on n variables was studied in [17].

Bent iterative constructions of self-dual Boolean bent functions in $n+2$ variables, based on concatenation of 4 Boolean bent functions in n variables, were presented in [3, 11]. In current work we give two constructions of generalized bent iterative functions that generalize the constructions for Boolean case:

Theorem 5. *1) Let f be a regular gbent function in n variables, then the sign function*

$$\left(F, \widetilde{F}, \widetilde{F}, -F\right),$$

is the sign function of a self-dual gbent function in $n+2$ variables;

2) Let f be a self-dual gbent function in n variables with the sign function F, and g be an anti-self-dual gbent function in n variables with the sign function G, then the sign function

$$(F, G, -G, F),$$

is the sign function of a gbent function in $n+2$ variables.

4 Sign functions of (anti-)self-dual gbent functions

Let I_n be the identity matrix of size n and $H_n = H_1^{\otimes n}$ be the n-fold tensor product of the matrix H_1 with itself, where

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

It is known the Hadamard property of this matrix

$$H_n H_n^T = 2^n I_{2^n},$$

where H_n^T is transpose of H_n (it holds $H_n^T = H_n$ by symmetricity of H_n).

Recall an orthogonal decomposition of $\mathbb{R}^{2^n}$ in eigenspaces of H_n from [3] (Lemma 5.2):

$$\mathbb{R}^{2^n} = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right) \oplus \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right),$$

where the symbol $\oplus$ denotes a direct sum of subspaces. Consider the same decomposition

$$\mathbb{C}^{2^n} = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right) \oplus \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right),$$

for a complex space $\mathbb{C}^{2^n}$.

As for the Boolean case (see [12]), we note that sign function of any self-dual gbent function is the eigenvector of $\mathcal{H}_n$ attached to the eigenvalue 1, that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n}) = \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$. The same holds for a sign function of any anti-self-dual gbent function, which obviously is an eigenvector of $\mathcal{H}_n$ attached to the eigenvalue (-1), that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}) = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$.

It is known that

$$\dim\left(\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})\right) = \dim\left(\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})\right) = 2^{n-1},$$

where $\dim(V)$ is the dimension of the subspace $V \subseteq \mathbb{R}^{2^n}$. Moreover, since $\mathcal{H}_n$ is symmetric (Hermitian), the subspaces $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n})$ and $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n})$ are mutually orthogonal.

In [11] it was proved that provided $n \geqslant 4$, the linear span of sign functions of self-dual as we;; as anti-self-dual Boolean bent functions in n variables has dimension 2^{n-1}. The same result can be also given for gbent functions:

Theorem 6. *Let $n \geqslant 4$, then the linear span of sign functions of (anti-)self-dual gbent functions in n variables has dimension 2^{n-1}.*

5 Self-dual gbent functions symmetric with respect to two variables

A generalized Boolean function $h \in \mathcal{GF}^q_{n+2}$ is symmetric with respect to two variables y and z if and only if there exist functions $f, g, s \in \mathcal{GF}^q_n$ such that

$$h(z, y, x) = f(x) + (y \oplus z)g(x) + yzs(x), \quad y, z \in \mathbb{F}_2, x \in \mathbb{F}_2^n. \tag{1}$$

In paper [23] it was proved that a function of such form is gbent if and only if the functions $f, f + g$ are gbent and $s(x) = q/2, x \in \mathbb{F}_2^n$. We study the conditions for self-duality of functions of such form.

Theorem 7. *Let h be a gbent function of the form (1). Then h is self-dual if and only if f is gbent, $g = \widetilde{f} + (q-1)f$, and $s(x) = q/2, x \in \mathbb{F}_2^n$.*

6 Affinity of self-dual gbent function

In paper [17] for the case when q is divisible by 4, necessary and sufficient conditions for the bentness of generalized Boolean functions of the form

$$f(x) = \sum_{i=1}^{n} \lambda_i x_i + \lambda_0,$$

where $\lambda_0, \lambda_1, ..., \lambda_n \in \mathbb{Z}_q$, were obtained. Functions from this class are referred to as *affine* functions.

It is well known that Boolean bent function and, as a consequence, self-dual Boolean bent function can not be affine. The next result shows non-existence of self-dual gbent functions in the class of affine functions.

Theorem 8. *There are no self-dual generalized bent functions in n variables of the form*

$$f(x) = \sum_{i=1}^{n} \lambda_i x_i + \lambda_0,$$

where $\lambda_0, \lambda_1, ..., \lambda_n \in \mathbb{Z}_q$.

7 Symmetries

Denote, according to [6], the orthogonal group of index n over the field $\mathbb{F}_2$ as

$$\mathcal{O}_n = \left\{ L \in GL\left(n, \mathbb{F}_2\right) | LL^T = I_n \right\},$$

where L^T denotes the transpose of L and I_n is an identical matrix of order n over the field $\mathbb{F}_2$.

In paper [5] (see also [3]) it was shown that the mapping

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x \rangle \oplus d,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$, preserves self-duality of a bent function. The group which consists of mappings of such form is called an *extended orthogonal group* and denoted by $\overline{\mathcal{O}}_n$ [4, 5]. It is known that this group is a subgroup of $\mathrm{GL}\left(n+2, \mathbb{F}_2\right)$ [5].

In paper [12] known results were generalized within isometric mappings from the set of all mappings of all Boolean functions in $n \geqslant 4$ variables into itself, which preserve the Hamming distance. Namely it was proved the necessity of such a form of mappings for prserving of (anti-)self-duality.

In current work we consider the mappings of the set of all generalized Boolean functions in n variables to itself of the form

$$f(x) \longrightarrow f(\pi(x)) + g(x),$$

where π is a permutation on the set $\mathbb{F}_2^n$ and $g \in \mathcal{GF}_n$. It is clear that such mappings preserve both Hamming and Lee distances between generalized Boolean functions.

The following result provides the construction of mappings of such form that preserves (anti-)self-duality of a Boolean function.

Theorem 9. *The mapping of the set of all generalized Boolean functions in n variables to itself of the form*

$$f(x) \longrightarrow f(\pi(x)) + g(x),$$

with

$$\pi(x) = L\left(x \oplus c\right),$$

and

$$g(x) = \frac{q}{2}\langle c, x \rangle + d,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$, preserves (anti-)self-duality of a gbent function.

References

[1] Canteaut A., Charpin P. Decomposing bent functions. *IEEE Trans. Inform. Theory*, **49**(8), 2004–2019 (2003).

[2] Carlet C. Boolean functions for cryptography and error correcting code. In: Crama Y., Hammer P.L. (eds.) Boolean Models and Methods in Mathematics, Computer Science, and Engineering. p. 257–397. Cambridge University Press, Cambridge (2010).

[3] Carlet C., Danielson L.E., Parker M.G., Solé. P. Self-dual bent functions. *Int. J. Inform. Coding Theory*, **1**, 384–399 (2010).

[4] Danielsen L.E., Parker M.G., Solé P. The Rayleigh quotient of bent functions. *Springer Lect. Notes in Comp. Sci.*, 5921, pp. 418–432. Springer, Berlin (2009).

[5] Feulner T., Sok L., Solé P., Wassermann A. Towards the Classification of Self-Dual Bent Functions in Eight Variables. *Des. Codes Cryptogr.*, **68**(1), 395–406 (2013).

[6] Janusz G.J. Parametrization of self-dual codes by orthogonal matrices. *Finite Fields Appl.*, **13**(3), 450–491 (2007).

[7] Hodžić S., Meidl W., Pasalic E. Full Characterization of Generalized Bent Functions as (Semi)-Bent Spaces, Their Dual, and the Gray Image. *IEEE Trans. Inform. Theory*, **64**(7), 5432–5440, 2018.

[8] Hodžić S., Pasalic E. Generalized Bent Functions — Some General Construction Methods and Related Necessary and Sufficient Conditions. *Cryptogr. Commun.*, **7**, 469–483, 2015.

[9] Hou X.-D. Classification of self dual quadratic bent functions. *Des. Codes Cryptogr.*, **63**(2), 183–198 (2012).

[10] Kutsenko A.V., *The Hamming Distance Spectrum Between Self-Dual Maiorana–McFarland Bent Functions*, Journal of Applied and Industrial Mathematics, **12**(1), 112–125 (2018).

[11] Kutsenko A. Metrical properties of self-dual bent functions. *Des. Codes Cryptogr.*, **88**, 201–222 (2020).

[12] Kutsenko A. The group of automorphisms of the set of self-dual bent functions. *Cryptogr. Commun.*, (2020). DOI: 10.1007/s12095-020-00438-y

[13] Martinsen T., Meidl W., Stănică P. Partial spread and vectorial generalized bent functions. *Des. Codes Cryptogr.*, **85**(1), 1–13 (2017).

[14] Mesnager S., Tang C., Qi Y., Wang L., Wu B., Feng K. Further Results on Generalized Bent Functions and Their Complete Characterization. *IEEE Trans. Inform. Theory*, 64(7), 4668–4674, 2018.

[15] Paterson K.G. Generalized Reed–Muller Codes and Power Control in OFDM Modulation. *IEEE Trans. Inform. Theory*, **46**(1), 104–120, 2000.

[16] Preneel B., Van Leekwijck W., Van Linden L., Govaerts R., Vandewalle J. Propagation characteristics of Boolean functions. In: Advances in Cryptology-EUROCRYPT. Lecture Notes in Computer Science, **473**, pp. 161–173. Springer, Berlin (1990).

[17] Singh B.K. On cross-correlation spectrum of generalized bent functions in generalized Maiorana–McFarland class. *Information Sciences Letters*, **2**(3), 139–145 (2013).

[18] Paterson K.G., Jones A.E. Efficient decoding algorithms for generalized Reed–Muller codes. *IEEE Trans. Commun.*, **48**(8), 1272–1285, 2000.

[19] Rothaus O.S. On bent functions. *J. Combin. Theory. Ser. A*, **20**(3). 300–305 (1976).

[20] Schmidt K.-U. Quaternary constant-amplitude codes for multicode CDMA. *IEEE Trans. Inform. Theory*, **55**, 1824–1832 (2009).

[21] Schmidt K.-U. $\mathbb{Z}_4$-valued quadratic forms and quaternary sequence families. *IEEE Trans. Inform. Theory*, **55**(10), 5803–5810 (2009).

[22] Sok L., Shi M., Solé P. Classification and Construction of quaternary self-dual bent functions. *Cryptogr. Commun.*, **10**(2), 277–289 (2018).

[23] Stănică P., Martinsen T., Gangopadhyay S., Singh B. K. Bent and generalized bent Boolean functions. *Des. Codes Cryptog.*, **69**, 77–94 (2013).

[24] Tang C., Xiang C., Qi Y., Feng K. Complete Characterization of Generalized Bent and 2^k-Bent Boolean Functions. *IEEE Trans. Inform. Theory*, **63**(7), 4668–4674, 2017.

[25] Tokareva N.N. Generalizations of bent functions — a survey. *J. Appl. Ind. Math.*, **5**(1), 110–129 (2011).

[26] Tokareva N.N. On the number of bent functions from iterative constructions: lower bounds. *Adv. Math. Commun.*, **5**(4), 609–621 (2011).

[27] Tokareva N. Bent Functions, Results and Applications to Cryptography. *Acad. Press. Elsevier*, 2015.

[28] Wada T. Characteristic bit sequences applicable to constant amplitude orthogonal multicode systems. *IEICE Trans. Fundamentals*, **E83-A**(11), 2160–2164, 2000.

Metrical properties of the set of bent functions in view of duality

Aleksandr Kutsenko and Natalia Tokareva

Abstract

In this work [1] we give a review of metrical properties of the entire set of bent functions and its significant subclasses of self-dual and anti-self-dual bent functions. We give results for iterative contruction of bent functions in $n+2$ variables based on the concatention of four bent functions and consider related open problem proposed by one of the authors. Criterion of self-duality for bent iterative functions and corollaries on sign functions and constructions of self-dual bent functions are discussed. It is explored that the pair of sets of bent functions and affine functions as well as a pair of sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables is a pair of mutually maximally distant sets that implies metrical duality. The solution to the problems of preserving bentness and anti-self-duality within automorphisms of the set of all Boolean functions is considered.

Keywords: Boolean bent function, self-dual bent function, Hamming distance, metrical regularity, automorphism group, iterative construction

1 Introduction

How much do we know about some cryptographic objects? One way to measure it is to describe what we can do with them. Otherwise to characterize groups of automorphisms of these objects — separately for each object or together while they form some special class. The question about the group of automorphisms of a set in the Boolean cube necessarily leads us to metrical properties of this set. That is why we are very interested in *metrical properties* of distinct cryptographic Boolean functions.

The term "bent function" was introduced by Oscar Rothaus in the 1960s [28]. It is known [36], that at the same time Boolean functions with maximal nonlinearity were also studied in the Soviet Union. The term

[1]The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

minimal function, which is actually a counterpart of a bent function, was proposed by the Soviet scientists Eliseev and Stepchenkov in 1962. Bent functions have connections with such combinatorial objects as Hadamard matrices and difference sets. Since bent functions have maximum Hamming distance to linear structures and affine functions they deserve attention for practical applications in symmetric cryptography, in particular, for block and stream ciphers. We refer to the survey [5] and monographies of Mesnager [25] and Tokareva [36] for more information concerning known results and open problems related to bent functions. Results regarding the study of metrical properties of the set of bent functions one can find in article [16].

In this paper we study the class of bent function $\mathcal{B}_n$ and its important subclasses — self-dual bent functions $\mathrm{SB}^+(n)$ (i.e. functions such that $f = \widetilde{f}$) and anti-self-dual bent functions $\mathrm{SB}^-(n)$ (i.e. functions such that $f \oplus 1 = \widetilde{f}$), where $\widetilde{f}$ is the dual of f. We suppose that the *keys* to the nontrivial and important properties of the class of bent functions are in understanding how does the *duality mapping* $f \to \widetilde{f}$ operate with bent functions. Recall that $\widetilde{\widetilde{f}} = f$ for every bent function f. It is important to note that the duality mapping is the *unique* known isometric mapping of the bent functions into themselves that can not be extended to a typical isometry of the whole set of all Boolean functions that preserves bent functions.

On other hand, the essence of bent functions is expressed in their metrical properties, namely in maximizing distances between them and affine functions. Note that this very idea in more general form is realized in the concept of metrical complement and metrically regular sets. Recall that $\widehat{X}$ is the metrical complement of the set of functions X if it contains all Boolean functions that are on the maximal possible distance from X. The set is metrically regular, if $\widehat{\widehat{X}} = X$. There is a some similarity to the self-duality of bent functions, is not it?

Our attention is drawn to automorphism groups of the sets $\mathcal{B}_n$, $\mathcal{A}_n$, $\mathrm{SB}^+(n)$, $\mathrm{SB}^-(n)$ and their metrical properties. Previously, we established that the set of all bent functions $\mathcal{B}_n$ and the set of all affine functions $\mathcal{A}_n$ form a pair of metrically regular sets, i.e. $\widehat{\widehat{\mathcal{B}_n}} = \widehat{\mathcal{A}_n} = \mathcal{B}_n$. Now we prove the same fact for the classes of self-dual and anti-self-dual functions: they form another such pair of metrically complement functions, i.e. $\widehat{\widehat{\mathrm{SB}^+(n)}} = \widehat{\mathrm{SB}^-(n)} = \mathrm{SB}^+(n)$. In both cases for elements in a pair of metrically regular sets we prove the coincidence of automorphism groups. Thus, $\mathrm{Aut}\,(\mathcal{B}_n) = \mathrm{Aut}\,(\mathcal{A}_n)$ and $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right)$. Some other curious properties of bent functions related to their special constructions are

discussed in the paper.

The work has the following structure: notation and definitions are in the Section 2. In Section 3 the duality of a bent function is described, including some its important properties and relevant hypothesis (Section 3.1). Some general and metrical properties of the set of bent functions which coincide with their duals, namely self-dual bent functions, are given in Section 3.2. In Section 4 we discuss the iterative construction of bent function in $n+2$ variables based on the concatenation of four bent functions in n variables. The lower bounds on its cardinality and open problem relevant for the set of bent function are in Section 4.1. Criterion of self-duality for bent iterative functions and its corollaries for sign functions together with constructions of self-dual bent functions are discussed in Sections 4.2 and 4.3. In Section 5 the metrical complement of the set of bent functions is studied (Section 5.2) and the results regarding metrical regularity of the set of bent functions and the set of affine functions are given. Metrical complement of the set of (anti-)self-dual bent functions is in Section 5.3. In Section 6 groups of automorphisms of considered sets are studied. The group of automorphisms of the set of bent functions is characterized in Section 6.3 while the (anti-)self-dual case is in Section 6.5. In Section 6.4 we discuss automorphisms of the set of all Boolean functions in n variables which define bijections between sets of self-dual and anti-self-dual bent functions. In Section 6.6 we state the relation between the results from Section 6.5 and preserving of the Rayleigh quotient of a Boolean function.

2 Notation

Let $\mathbb{F}_2^n$ be a space of binary vectors of length n. Denote, following [12], the orthogonal group of index n over the field $\mathbb{F}_2$ as

$$\mathcal{O}_n = \left\{ L \in GL\left(n, \mathbb{F}_2\right) | LL^T = I_n \right\},$$

where L^T denotes the transpose of L and I_n is an identical matrix of order n over the field $\mathbb{F}_2$.

A *Boolean function* f in n variables is a map from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. Its *sign function* is $F(x) = (-1)^{f(x)}$, $x \in \mathbb{F}_2^n$. We will also refer to a sign function as to a vector from the set $\{\pm 1\}^{2^n}$:

$$F = (-1)^f = \left((-1)^{f_0}, (-1)^{f_1}, ..., (-1)^{f_{2^n-1}}\right) \in \{\pm 1\}^{2^n},$$

where $(f_0, f_1, ..., f_{2^n-1}) \in \mathbb{F}_2^{2^n}$ is a truth-table representation of f with arguments given in the lexicographic order. The set of Boolean functions in n variables is denoted by $\mathcal{F}_n$.

The *algebraic normal form* (ANF, Zhegalkin polynomial) of a Boolean function $f \in \mathcal{F}_n$ is defined to be

$$f(x_1, x_2, ..., x_n) = \bigoplus_{(i_1,i_2,...,i_n)\in\mathbb{F}_2^n} a_{i_1 i_2 ... i_n} x_1^{i_1} x_2^{i_2} ... x_n^{i_n},$$

where $a_z \in \mathbb{F}_2$ for any $z \in \mathbb{F}_2^n$ (with the convention $0^0 = 1$). The *algebraic degree* $\deg(f)$ of a Boolean function f is the maximal degree of monomials which occur in its algebraic normal form with nonzero coefficients.

The *Hamming weight* $\mathrm{wt}(x)$ of the vector $x \in \mathbb{F}_2^n$ is the number of nonzero coordinates of x. The *Hamming weight* $\mathrm{wt}(f)$ of the function $f \in \mathcal{F}_n$ is the Hamming weight of its vector of values. The sign $\oplus$ denotes a sum modulo 2. The *Hamming distance* $\mathrm{dist}(f, g)$ between Boolean functions f, g in n variables is a cardinality of the set $\{x \in \mathbb{F}_2^n : f(x) \oplus g(x) = 1\}$. For $x, y \in \mathbb{F}_2^n$ denote $\langle x, y\rangle = \bigoplus_{i=1}^{n} x_i y_i$. Boolean functions in n variables of the form $f(x) = \langle a, x\rangle \oplus a_0, x \in \mathbb{F}_2^n$, where $a_0 \in \mathbb{F}_2, a \in \mathbb{F}_2^n$, are called *affine* functions. The set of affine functions in n variables is denoted by $\mathcal{A}_n$.

The *Walsh–Hadamard transform* (WHT) of a Boolean function f in n variables is an integer valued function $W_f : \mathbb{F}_2^n \to \mathbb{Z}$, defined as

$$W_f(y) = \sum_{x\in\mathbb{F}_2^n} (-1)^{f(x)\oplus\langle x,y\rangle}, \quad y \in \mathbb{F}_2^n.$$

A Boolean function f in an even number n of variables is called *bent* if

$$|W_f(y)| = 2^{n/2},$$

for all $y \in \mathbb{F}_2^n$. The set of all bent functions in n variables is denoted by $\mathcal{B}_n$.

3 The dual of a bent function

From the definition of a bent function it follows that for any $y \in \mathbb{F}_2^n$ we have

$$W_f(y) = (-1)^{\widetilde{f}(y)} 2^{n/2},$$

for some $\widetilde{f} \in \mathcal{F}_n$. The Boolean function $\widetilde{f}$ defined above is called the *dual* function of the bent function f. Thus, for any bent function in n variables its dual Boolean function is uniquely defined. The duality of bent functions was introduced by Dillon [10].

3.1 Properties

Some basic known properties of dual functions are the following [5]:

- Every dual function is a bent function;
- If $\widetilde{f}$ is dual to f and $\widetilde{\widetilde{f}}$ is dual to $\widetilde{f}$, then $\widetilde{\widetilde{f}} = f$;
- The mapping $f \to \widetilde{f}$ which acts on the set of bent functions, preserves the Hamming distance.

There is the following connection between the algebraic degrees of a bent function and its dual [13]:

$$n/2 - \deg(f) \geqslant \frac{n/2 - \deg\left(\widetilde{f}\right)}{\deg\left(\widetilde{f}\right) - 1}.$$

Some results obtained for dual functions can be used in proving the results concerning bent functions, in particular, the connection between algebraic normal form (ANF) coefficients of a bent function and its dual, see [7]:

$$\sum_{x \preccurlyeq y} f(x) = 2^{\mathrm{wt}(y)} - 2^{n/2-1} + 2^{\mathrm{wt}(y)-n/2} \sum_{x \preccurlyeq y \oplus 1} \widetilde{f}(x).$$

One of the most important problem in bent functions is to find the number of them. A new approach to this problem was introduced in [32], see Section 4.1, and the following hypothesis was formulated.

Hypothesis (Tokareva, 2011): *any Boolean function in n variables of degree not more than $n/2$ can be represented as the sum of two bent functions in n variables, where $n \geqslant 2$ is an even number.*

The review of partial results regarding this problem and also in favour of the Hypothesis one can find in [34]. It was also proved in [35] that

Theorem 1. *A bent function in $n \geqslant 4$ variables can be represented as the sum of two bent functions in n variables if and only if its dual bent function does.*

So, it follows that the mentioned Hypothesis with the decomposition problem, see Section 4.1, can not be considered separately for a bent function and its dual.

It is worth noting that this hypothesis is a counterpart of the Goldbach's conjecture in number theory unsolved since 1742: any even number $n > 4$ can be represented as the sum of two prime numbers.

Isometric mappings of the set of all Boolean functions in n variables to itself which preserve bentness and the Hamming distance between every bent function and its dual were characterized in [19], namely it was proved that

Theorem 2. *An isometric mapping φ of the set of all Boolean functions in n variables into itself preserves bentness and the Hamming distance between every bent function and its dual if and only if φ has form*

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x\rangle \oplus d, \quad x \in \mathbb{F}_2^n,$$

for some $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$.

3.2 Self-duality

If a bent function f coincides with its dual it is said to be *self-dual*, that is $f = \widetilde{f}$. A bent function which coincides with the negation of its dual is called an *anti-self-dual*, that is $f = \widetilde{f} \oplus 1$. The set of (anti-)self-dual bent functions in n variables, according to [14], is denoted by $\mathrm{SB}^+(n)$ $\left(\mathrm{SB}^-(n)\right)$.

Self-dual bent functions were explored in paper of Carlet et. al. [4] in 2010, where some important properties and constructions were given. All equivalence classes of self-dual bent functions in 2, 4, and 6 variables and all quadratic self-dual bent functions in 8 variables with respect to a restricted form of an affine transformation

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x\rangle \oplus d, \quad x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$, which preserves self-duality were also presented. Further, equivalence classes of cubic self-dual bent functions in 8 variables with respect to the mentioned above restricted form of affine transformation one can find in [11]. In [14] a classification of quadratic self-dual bent functions was obtained. The upper bound for the cardinality of the set of self-dual bent functions was given in [15]. In [20, 24] one can find new constructions of self-dual bent functions. A connection of quaternary self-dual bent functions and self-dual bent Boolean functions was shown in [29]. In [18] it was proved that for any $d \in \{2, 3, ..., n/2\}$ there exists a self-dual bent function of algebraic degree d.

In papers [17, 18, 19] metrical properties of the sets of (anti-)self-dual bent functions in n variables were studied. Below we briefly discuss some of them.

Recall that bent functions in $2k$ variables which have a representation

$$f(x, y) = \langle x, \pi(y)\rangle \oplus g(y), \quad x, y \in \mathbb{F}_2^k,$$

where $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and g is a Boolean function in k variables, form the well known *Maiorana–McFarland class* of bent functions [23]. Let the denotion $\mathrm{SB}^+_{\mathcal{M}}(n)$ stands for the set of self-dual Maiorana–McFarland bent functions and $\mathrm{SB}^-_{\mathcal{M}}(n)$ for the set of anti-self-dual ones. Necessary and sufficient conditions of (anti-)self-duality of bent functions from Maiorana–McFarland class are known from [4]. Regarding the spectrum of Hamming distances in [17] the following result was proved.

Theorem 3. *Let* $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n) \cup \mathrm{SB}^-_{\mathcal{M}}(n)$*, then*

$$\mathrm{dist}(f,g) \in \left\{ 2^{n-1}, 2^{n-1}\left(1 \pm \frac{1}{2^r}\right), r = 0, 1, ..., n/2 - 1 \right\},$$

Moreover, if either $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n)$ *or* $f, g \in \mathrm{SB}^-_{\mathcal{M}}(n)$*, then all distances except* 2^{n-1} *are attainable, and for any pair* $f \in \mathrm{SB}^+_{\mathcal{M}}(n)$ *and* $g \in \mathrm{SB}^-_{\mathcal{M}}(n)$ *it holds* $\mathrm{dist}(f,g) = 2^{n-1}$*.*

By analysis of the set of distances from Theorem 3 the minimal Hamming distance between considered functions can be obtained:

Corollary 1. *The minimal Hamming distance between (anti-)self-dual Maiorana–McFarland bent functions is equal to* 2^{n-2}*.*

Moreover, since the minimal Hamming distance between quadratic Boolean functions in n variables (which correspond to codewords of the $\mathrm{RM}(2,n)$ code) is at least 2^{n-2} [21], the following fact holds

Corollary 2. *The minimal Hamming distance between quadratic bent functions can be attained on (anti-)self-dual Maiorana–McFarland bent functions.*

It is well known that the minimal Hamming distance between bent functions in n variables is equal to $2^{n/2}$, see [16] for instance. In [18] it was proved that this extremal value can be attained on (anti-)self-dual bent functions.

Theorem 4. *Let* $n \geqslant 4$*, then the minimal Hamming distance between distinct (anti-)self-dual bent functions in* n *variables is equal to* $2^{n/2}$*.*

4 Iterative construction $\mathcal{BI}$

Let f_0, f_1, f_2, f_3 be Boolean functions in n variables. Consider a Boolean function g in $n+2$ variables which is defined as

$$g(00,x) = f_0(x), \quad g(01,x) = f_1(x),$$

$$g(10, x) = f_2(x), \quad g(11, x) = f_3(x),$$

where $x \in \mathbb{F}_2^n$.

It is known (Preneel et. al., 1991; see also [1, 32]) that under condition $f_0, f_1, f_2, f_3 \in \mathcal{B}_n$ the mentioned function g is a bent function in $n+2$ variables if and only if

$$\widetilde{f}_0 \oplus \widetilde{f}_1 \oplus \widetilde{f}_2 \oplus \widetilde{f}_3 = 1,$$

that gives the construction of a bent function in $n+2$ variables through the concatenation of vectors of values of four bent functions in n variables [27].

Bent functions which are obtained by this construction, in accordance with [32], are called *bent iterative functions* ($\mathcal{BI}$) and the set of such bent functions in n variables is denoted by $\mathcal{BI}_n$.

In the article [6] the comparison of cardinalities of different known iterative constructions of bent functions in $n \leqslant 10$ variables was presented and the class $\mathcal{BI}$ had the biggest cardinality among them.

According to [1] there exist bent functions from Maiorana–McFarland class [23] and from the class $\mathcal{PS}$ (Partial Spreads) [10] that can not be represented as bent iterative functions. Also from paper [2] on nonnormal bent functions it follows that there exist bent functions in $\mathcal{BI}_n$ that are nonequivalent to Maiorana–McFarland bent functions.

4.1 Lower bounds on the cardinality and related open problem

In paper [32] some possible ways of how to calculate the number of bent iterative functions were shown.

Theorem 5. *For any even* $n \geqslant 4$

$$|\mathcal{BI}_n| = \sum_{f' \in \mathcal{B}_{n-2}} \sum_{f'' \in \mathcal{B}_{n-2}} |(\mathcal{B}_{n-2} \oplus f') \cap (\mathcal{B}_{n-2} \oplus f'')| .$$

Denote $X_n = \{f \oplus h | f, h \in \mathcal{B}_n\}$ and consider the system $\{C_f : f \in \mathcal{B}_n\}$ of its subsets defined as $C_f = \mathcal{B}_n \oplus f$. So,

$$X_n = \bigcup_{f \in \mathcal{B}_n} C_f.$$

Let ψ be an element of X_n. The number of subsets C_f that cover ψ, according to [32], is called *multiplicity* of ψ and is denoted by $m(\psi)$. One can notice that if ψ is covered by C_f then it is covered by any set $C_{f'}$, where f' is obtained from f by adding an affine function.

In [32] the exact number of bent iterative functions through the multiplicities was obtained.

Theorem 6. *For any even* $n \geqslant 2$

$$|\mathcal{BI}_{n+2}| = \sum_{\psi \in C_f} m^2(\psi).$$

So, in order to evaluate $|\mathcal{BI}_{n+2}|$ (and then $|\mathcal{B}_{n+2}|$) we have to study the set X_n and the distribution of multiplicities for its elements. Such analysis, as shown in [32], gives the following lower bound.

Theorem 7. *For any even* $n \geqslant 2$

$$\frac{|\mathcal{B}_{n+2}|^4}{|X_n|} \leqslant |\mathcal{BI}_{n+2}| \leqslant |\mathcal{B}_{n+2}|.$$

Thus for calculating the exact number of bent iterative functions one has to study the structure of the set X_n. So, we come to a new problem statement.

Open problem: bent sum decomposition (Tokareva, 2011). *What Boolean functions can be represented as the sum of two bent functions in n variables? How many such representations does a Boolean function admit?*

The related Hypothesis was previuosly mentioned in the Section 3.1.

4.2 Self-dual bent iterative functions

The set of (anti-)self-dual bent functions from $\mathcal{BI}_n$ is further denoted by $\mathrm{SB}^+_{\mathcal{BI}}(n)$ $\left(\mathrm{SB}^-_{\mathcal{BI}}(n)\right)$.

In paper [18] the necessary and sufficient conditions of self-duality of bent iterative functions were studied, namely the following result was obtained.

Theorem 8. *Let* $g \in \mathcal{BI}_{n+2}$ *then* g *is self-dual if and only if there exists such pair of functions* $g_1, g_2 \in \mathcal{B}_n$ *and a function* $h \in \mathcal{F}_n$ *that:*

$$\begin{aligned}
f_0 &= (g_1 \oplus g_2)\, h \oplus g_1 = \widetilde{g}_2,\\
f_1 &= (g_1 \oplus g_2)\, h \oplus g_2 = \widetilde{g_1 \oplus h},\\
f_2 &= (g_1 \oplus g_2)\, h \oplus g_2 \oplus h = \widetilde{g}_1,\\
f_3 &= (g_1 \oplus g_2)\, h \oplus g_1 \oplus h \oplus 1 = \widetilde{g_2 \oplus h} \oplus 1.
\end{aligned}$$

Remark 1. *It can be proved that the function* h *is uniquely defined by a pair of bent functions* g_1, g_2*, namely:* $h = g_1 \oplus \widetilde{g}_1 \oplus g_2 \oplus \widetilde{g}_2$.

By considering constant function h one can immediately obtain two constructions of self-dual bent iterative functions.

Corollary 3. *Functions*

$$f'(y_1, y_2, x) = (y_1 \oplus y_2)\left(f(x) \oplus \widetilde{f}(x)\right) \oplus f(x) \oplus y_1 y_2,$$

$$f''(y_1, y_2, x) = (y_1 \oplus y_2)(\varphi(x) \oplus \omega(x)) \oplus \varphi(x) \oplus \alpha_1 y_1 \oplus \alpha_2 y_2 \oplus y_1 y_2,$$

where

$$y_1, y_2, \alpha_1, \alpha_2 \in \mathbb{F}_2, \alpha_1 \oplus \alpha_2 = 1, x \in \mathbb{F}_2^n,$$
$$f \in \mathcal{B}_n, \varphi \in \mathrm{SB}^+(n), \omega \in \mathrm{SB}^-(n),$$

are self-dual bent functions in $n+2$ variables.

Remark 2. *The first construction from those listed above (for f') was presented in [4] as an example of the construction which uses the indirect sum of bent functions, see [3]. It is worth noting that the second construction (for f'') can also be obtained from indirect sum of bent functions.*

Since these constructions do not intersect, the sum of their cardinalities is a lower bound for the cardinality of the set of self-dual bent iterative functions.

Corollary 4. *It holds*

$$|\mathcal{B}_{n-2}| + \left|\mathrm{SB}^+(n-2)\right|^2 \leqslant \left|\mathrm{SB}^+_{\mathcal{BI}}(n)\right| \leqslant |\mathcal{B}_{n-2}|^2.$$

4.3 The dimension of linear span of sign functions of self-dual bent functions

Let I_n be an identity matrix of size n and $H_n = H_1^{\otimes n}$ be the n-fold tensor product of the matrix H_1 with itself, where

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

It is known the Hadamard property of this matrix

$$H_n H_n^T = 2^n I_{2^n}.$$

Denote $\mathcal{H}_n = 2^{-n/2} H_n$. In terms of sign functions the function $f \in \mathcal{F}_n$ is bent if for its sign function F it holds $\mathcal{H}_n F \in \{\pm 1\}^{2^n}$.

Recall that a non-zero vector $v \in \mathbb{C}^n$ is called an *eigenvector* of a square $n \times n$ matrix A attached to the eigenvalue $\lambda \in \mathbb{C}$ if $Av = \lambda v$. A linear span of eigenvectors attached to the eigenvalue λ is called an *eigenspace* associated

with λ. Consider a linear mapping $\psi : \mathbb{C}^n \to \mathbb{C}^n$ represented by a $n \times n$ complex matrix A. A *kernel* of ψ is the set

$$\mathrm{Ker}\,(\psi) = \{x \in \mathbb{C}^n | Ax = \mathbf{0} \in \mathbb{C}^n\},$$

where $\mathbf{0}$ is a zero element of the space $\mathbb{C}^n$.

From the definition of self-duality it follows that sign function of any self-dual bent function is the eigenvector of $\mathcal{H}_n$ attached to the eigenvalue 1, that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n - I_{2^n}) = \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$. The same holds for a sign function of any anti-self-dual bent function, which obviously is an eigenvector of $\mathcal{H}_n$ attached to the eigenvalue (-1), that is an element from the subspace $\mathrm{Ker}\,(\mathcal{H}_n + I_{2^n}) = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$.

In [4] an orthogonal decomposition of $\mathbb{R}^{2^n}$ in eigenspaces of H_n was given:

$$\mathbb{R}^{2^n} = \mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right) \oplus \mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right), \qquad (1)$$

where the symbol $\oplus$ denotes a direct sum of subspaces.

It is known that

$$\dim\left(\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)\right) = \dim\left(\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)\right) = 2^{n-1},$$

where $\dim(V)$ is the dimension of the subspace $V \subseteq \mathbb{R}^{2^n}$. Moreover, from symmetricity of $\mathcal{H}_n$ it follows that the subspaces $\mathrm{Ker}\left(H_n - 2^{n/2} I_{2^n}\right)$ and $\mathrm{Ker}\left(H_n + 2^{n/2} I_{2^n}\right)$ are mutually orthogonal.

In [18] it was proved that within the set of sign functions of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables there exist basises of the eigenspaces of the matrix $\mathcal{H}_n$ attached to the eigenvalues 1 and (-1) correspondingly.

Theorem 9. *The linear span of sign functions of (anti-)self-dual bent functions in $n \geqslant 4$ variables has dimension 2^{n-1}.*

It is worth notice that the desired basises consist of sign functions of (anti-)self-dual bent iterative functions provided by two constructions from Corollary 3.

5 Metrical complement and regularity

In this section we give results regarding notable metrical property of a subset of Boolean cube called metrical regularity. The sets of affine Boolean functions and bent functions possess it. The sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables are also mutually maximaly distant.

That implies metrical *duality*, in some sence, between the considered pairs of subsets of Boolean functions.

Regarding that some essential and intriguing questions arise: for instance, are there any pairs of metrically regular subsets inside the metrically regular set of bent functions in n variables? If additionally, in order to exclude some trivial cases we consider only the subsets which include functions together with their negations, the maximal Hamming distance from the considered sets is at most 2^{n-1}. Are there any pairs of metrically regular subsets with additional mentioned requirement such that the distance between them is exactly 2^{n-1}, that is to say they are extremal in a manner?

5.1 Definitions

Let $X \subseteq \mathbb{F}_2^n$ be an arbitrary set and let $y \in \mathbb{F}_2^n$ be an arbitrary vector. Define the *distance* between y and X as $\mathrm{dist}(y, X) = \min\limits_{x \in X} \mathrm{dist}(y, x)$. The *maximal distance* from the set X is

$$\mathrm{d}(X) = \max_{y \in \mathbb{F}_2^n} \mathrm{dist}(y, X).$$

In coding theory this number is also known as the *covering radius* of the set X. A vector $z \in \mathbb{F}_2^n$ is called *maximally distant* from a set X if $\mathrm{dist}(z, X) = \mathrm{d}(X)$. The set of all maximally distant vectors from the set X is called the *metrical complement* of the set X and denoted by $\widehat{X}$. A set X is said to be *metrically regular* if $\widehat{\widehat{X}} = X$. Define, a subset of Boolean functions to be *metrically regular* if the set of corresponding vectors of values is metrically regular [36].

Sets of functions which have maximum distance from partition set functions were studied in [30], it was shown that partition set functions defined by some partition are mutually maximally distant sets. Lower bound on size of the largest metrically regular subset of the Boolean cube was studied in [26].

5.2 The set of bent functions

It is well-known that

Proposition 1. *Any isometric mapping of the form*

$$f(x) \longrightarrow f\left(Ax \oplus b\right) \oplus \langle c, x \rangle \oplus d,$$

where $A \in \mathrm{GL}(n)$, $b, c \in \mathbb{F}_2^n$, $d \in \mathbb{F}_2$, *preserves bentness.*

In [33] the following theorem was proved:

Theorem 10. *For each non-affine Boolean function $h \in \mathcal{F}_n$ there exists a bent function $f \in \mathcal{B}_n$ such that $f \oplus h$ is not bent.*

From Proposition 1 and Theorem 10 it follows that the set of bent functions is closed under addition of affine Boolean functions only. This fact implies that the affine functions are precisely all Boolean functions which are at the maximum distance from the class of bent functions. Namely, in [33] it was shown that

Theorem 11. *A Boolean function in n variables is*

- *a bent function if and only if it has the maximal possible distance $2^{n-1} - 2^{n/2-1}$ to the set of all affine functions, that is it is an element of $\widehat{\mathcal{A}_n}$;*
- *an affine function if and only if it has the maximal possible distance $2^{n-1} - 2^{n/2-1}$ to the set of all bent functions, that is it is an element of $\widehat{\mathcal{B}_n}$.*

Thus, from the results given in [33] it follows that there exists a *duality*, in some sense, between the definitions of bent functions and affine functions. In particular, we obtain metrical regularity of the sets of affine functions and bent functions.

Corollary 5. *It holds:*

- *the set $\mathcal{A}_n$ of all affine Boolean functions in n variables is metrically regular;*
- *the set $\mathcal{B}_n$ of all bent functions in n variables is metrically regular.*

5.3 The set of (anti-)self-dual bent functions

Since for any self-dual Boolean function $f \in \mathrm{SB}^+(n)$ its negation $f \oplus 1$ is also self-dual, the maximal Hamming distance from the set $\mathrm{SB}^+(n)$ is at most 2^{n-1}. It was proved by Carlet et. al. in [4] that the Hamming distance between any pair of self-dual and anti-self-dual bent functions, both in n variables, is equal to 2^{n-1}. From that it follows that

$$\mathrm{d}\left(\mathrm{SB}^+(n)\right) = 2^{n-1},$$

and all anti-self-dual bent functions in n variables belong to the metrical complement of the set of self-dual bent functions in n variables.

In paper [18] the metrical complement of the set of (anti-)self-dual bent functions in $n \geqslant 4$ variables was completely characterized by using the orthogonal decomposition (1) and the existence of the basis provided by the Theorem 9, namely it was proven that

Theorem 12. *Let $n \geqslant 4$, then the following statements hold:*

- *The metrical complement of the set of self-dual bent functions coincides with the set of anti-self-dual bent functions;*
- *The metrical complement of the set of anti-self-dual bent functions coincides with the set of self-dual bent functions.*

As for the pair of the sets of bent functions and affine functions, it follows that there exists a *duality*, in some sense, between the sets of self-dual and anti-self-dual bent functions in $n \geqslant 4$ variables.

The case $n = 2$ was considered explicitely and it appeared that both $\mathrm{SB}^+(2)$ and $\mathrm{SB}^-(2)$ are metrically regular sets. From that and the Theorem 12 it follows

Theorem 13. *The sets $\mathrm{SB}^+(n)$, $\mathrm{SB}^-(n)$ are metrically regular sets, both with covering radius 2^{n-1}.*

6 The group of automorphisms

Study of automorphism groups of mathematical objects deserves attention since these groups are closely connected with the structure of the objects. There exists a natural question: how groups of automorphisms of two mathematical objects, one of which is embedded to another one, are related.

An example of such a problem statement is the set of bent functions in n variables and one of its significant subclasses which consisits of self-dual bent functions in n variables.

It is also worth mentioning that the complexity of classification of combinatorial objects depends on generality of the approach. Consequently, the question '*if the common approach to classify (self-dual) bent functions is the most general within automorphisms of the set of Boolean functions*', arises naturally.

6.1 Isometric mappings and automorphism groups

A mapping φ of the set of all Boolean functions in n variables to itself is called *isometric* if it preserves the Hamming distance between functions, that is

$$\mathrm{dist}(\varphi(f), \varphi(g)) = \mathrm{dist}(f, g),$$

for any $f, g \in \mathcal{F}_n$. Following [19] denote the set of all isometric mappings of the set of all Boolean functions in n variables to itself by $\mathcal{I}_n$.

It is known (A. A. Markov, 1956) that every isometric mapping of all Boolean functions in n variables to itself has the unique representation of the form

$$f(x) \longrightarrow f(\pi(x)) \oplus g(x), \tag{2}$$

where π is a permutation on the set $\mathbb{F}_2^n$ and $g \in \mathcal{F}_n$ [22]. The mapping of this form is denoted by $\varphi_{\pi,g} \in \mathcal{I}_n$.

The *group of automorphisms* of a fixed subset $M \subseteq \mathcal{F}_n$ is the group of isometric mappings of the set of all Boolean functions in n variables to itself preserving the set M. It is denoted by $\mathrm{Aut}\,(M)$.

6.2 Matrix representation

For a number $k \in \{0, 1, ..., 2^n - 1\}$ denote by $\mathbf{v}_k \in \mathbb{F}_2^n$ its binary representation.

Recall that a square matrix is called *monomial* (or *generalized permutation matrix*) if it has exactly one nonzero entry in each row and each column.

There is an one-to-one correspondence between the set $\mathcal{I}_n$ and the set of monomial matrices of order $2^n \times 2^n$ with nonzero elements from the set $\{\pm 1\}$. Indeed, consider an arbitrary mapping $\varphi_{\pi,g} \in \mathcal{I}_n$. Then for any $f \in \mathcal{F}_n$ and its sign function

$$F = \left((-1)^{f(\mathbf{v}_0)}, (-1)^{f(\mathbf{v}_1)}, ..., (-1)^{f(\mathbf{v}_{2^n-1})}\right) \in \{\pm 1\}^{2^n},$$

the sign function

$$F' = \left((-1)^{f'(\mathbf{v}_0)}, (-1)^{f'(\mathbf{v}_1)}, ..., (-1)^{f'(\mathbf{v}_{2^n-1})}\right) \in \{\pm 1\}^{2^n},$$

of $f' = \varphi_{\pi,g}(f) \in \mathcal{F}_n$ can be expressed as $F' = AF$, where A is a $2^n \times 2^n$ monomial matrix, constructed by the permutation π and the function g:

$$i \begin{pmatrix} & & & \overset{j}{\vdots} & & & \\ & & & 0 & & & \\ & & & \vdots & & & \\ \dots & 0 & \dots & (-1)^{g(\mathbf{v}_{i-1})} & \dots & 0 & \dots \\ & & & \vdots & & & \\ & & & 0 & & & \\ & & & \vdots & & & \end{pmatrix},$$

in which in the i-th row a nonzero element $(-1)^{g(\mathbf{v}_{i-1})}$ is in the j-th column, where $(j-1)$ is a number with binary representation $\pi(\mathbf{v}_{i-1})$. So the i-th

component of $F' = AF$ is equal to

$$(-1)^{f'(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))} \cdot (-1)^{g(\mathbf{v}_{i-1})} = (-1)^{f(\pi(\mathbf{v}_{i-1}))\oplus g(\mathbf{v}_{i-1})},$$

for any $i \in \{1, 2, ..., 2^n\}$, that is equivalent to

$$f'(x) = f(\pi(x)) \oplus g(x), \quad x \in \mathbb{F}_2^n.$$

6.3 The group of automorphisms of the set of bent functions

Some attempts to determine the automorphism group of a given bent function were undertaken by Dempwolff [9] in 2006. Results were presented in terms of elementary Abelian Hadamard difference sets (equivalently, bent functions).

A natural question whether there exist isometric mappings of Boolean functions into itself, distinct from those mentioned in Proposition 1, which preserve the class of bent function was completely solved in paper [31], where it was proved that there were no other mappings possessing such a property. Namely by using the Theorem 11 in view of the duality the following coincidence was shown.

Theorem 14.

$$\mathrm{Aut}(\mathcal{B}_n) = \mathrm{Aut}(\mathcal{A}_n).$$

Note that the set of all affine functions in n variables forms a group isomorphic to $\mathbb{F}_2^{n+1}$. The group of automorphisms of the set of all affine functions in n variables consists, as it is well known, of mappings of the form (2) with affine permutation π and affine shift g, see, for example, [21]. So, the result is formulated as follows.

Theorem 15. *It holds*

$$\mathrm{Aut}(\mathcal{B}_n) = \mathrm{GA}(n) \ltimes \mathbb{F}_2^{n+1},$$

where $\mathrm{GA}(n)$ *stands for the affine group and the symbol* $\ltimes$ *for semidirect product.*

These results imply the non-existence of a more general approach to equivalence of bent functions than that on the base of isometric mappings.

6.4 Isometric bijections between self-dual and anti-self-dual bent functions

It is known [4] that there exists a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$, based on the decomposition of sign functions of (anti-)self-dual bent functions. Also note that from the existence of such bijection it follows that $\left|\mathrm{SB}^+(n)\right| = \left|\mathrm{SB}^-(n)\right|$.

Namely, let $(Y, Z) \in \{\pm 1\}^{2^n}$, where $Y, Z \in \{\pm 1\}^{2^{n-1}}$, be a sign function for some $f \in \mathrm{SB}^+(n)$. Then a vector $(Z, -Y) \in \{\pm 1\}^{2^n}$ is a sign function for some function from $\mathrm{SB}^-(n)$. In terms of isometric mappings the mentioned transformation can be represented as

$$f(x) \longrightarrow f\left(x \oplus c\right) \oplus \langle c, x \rangle ,$$

where $c = (1, 0, 0, ..., 0) \in \mathbb{F}_2^n$.

In paper [14] it was mentioned that the more general form of this mapping

$$f(x) \longrightarrow f\left(x \oplus c\right) \oplus \langle c, x \rangle ,$$

where $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$. It is obvious that this mapping is an element from $\mathcal{I}_n$.

In paper [19] these results were generalized within isometric mappings from the set $\mathcal{I}_n$ for $n \geqslant 4$.

The criterion of bijectivity between self-dual and anti-self-dual bent functions was obtained in [19] with a use of the orthogonal decomposition (1) and the basis from the Theorem 9.

Theorem 16. *Let $n \geqslant 4$, then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ with matrix A is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if $A\mathcal{H}_n = -\mathcal{H}_n A$.*

By using this criterion in [19] the general form of considered isometric bijections was found.

Theorem 17. *For $n \geqslant 4$ isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ if and only if*

$$\pi(x) = L\left(x \oplus c\right), \ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is odd, $d \in \mathbb{F}_2$.

6.5 The group of automorphisms of the set of (anti-)self-dual bent functions

In [4] the followng problem was pointed:

Open question (Carlet, Danielson, Parker, Solé, 2010): *to find mappings preserving self-duality, distinct from the known ones, or give a proof that there are no more.*

In paper [19] this question was resolved within isometric mappings of the set of all Boolean functions in $n \geqslant 4$ variables into itself.

At first the problem of how the sets of isometric mapping preserving self-duality and anti-self-duality or, in other words, groups of automorphisms of the sets $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$ are related. This problem was solved in [19], where with a use of the orthogonal decomposition (1) and the basis from the Theorem 9, the criterion of preserving self-duality was given.

Theorem 18. *Let $n \geqslant 4$, then for isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ with matrix A the following conditions are equivalent:*

1) $\varphi_{\pi,g}$ preserves self-duality;

2) $\varphi_{\pi,g}$ preserves anti-self-duality;

3) $A\mathcal{H}_n = \mathcal{H}_n A$.

From this result it follows that

Corollary 6. *For $n \geqslant 4$ it holds* $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right)$.

The problem of characterizing mappings which preserve self-duality was studied by Carlet et. al. in [4] and Feulner et. al. in [11], where it was shown that the mapping

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x\rangle \oplus d,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$, preserves self-duality of a bent function. It is obvious that this mapping is isometric and corresponds to $\varphi_{\pi,g} \in \mathcal{I}_n$ with

$$\pi(x) = L\left(x \oplus c\right),\ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \langle c, x\rangle \oplus d,\ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$. The group which consists of mappings of such form is called an *extended orthogonal group* and denoted by $\overline{\mathcal{O}}_n$ [8, 11]. It is known that this group is a subgroup of $\mathrm{GL}\left(n+2, \mathbb{F}_2\right)$ [11].

In paper [19] known results were generalized within isometric mappings from the set $\mathcal{I}_n$ for $n \geqslant 4$. Namely by using the criterion from Theorem 18 and the matrix representation of isometric mappings it was obtained that the desired group of automorphisms coincides with the extended orthogonal group.

Theorem 19. *For $n \geqslant 4$ it holds*

$$\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right) = \overline{\mathcal{O}}_n.$$

In view of Theorems 17 and 19 it appeares that bijections and mappings which preserve self-duality are quite similar except the parity of the vector $c \in \mathbb{F}_2^n$, which 'switches' them in some sence.

It follows that the classification of self-dual bent functions in $n \geqslant 4$ variables based on the restricted form of affine equivalence proposed in articles [4, 11] is the most general within isometric mappings of the set of all Boolean functions in n variables into itself.

6.6 Isometric mappings and the Rayleigh quotient

In [4] the *Rayleigh quotient* S_f of a Boolean function $f \in \mathcal{F}_n$ was defined as

$$S_f = \sum_{x,y \in \mathbb{F}_2^n} (-1)^{f(x) \oplus f(y) \oplus \langle x,y \rangle} = \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)} W_f(y).$$

In a scope of bent functions the Rayleigh quotient characterizes the Hamming distance between a bent function and its dual. Indeed, let $f \in \mathcal{B}_n$, then

$$\mathrm{dist}(f, \widetilde{f}) = 2^{n-1} - \frac{1}{2^{n/2+1}} S_f = 2^{n-1} - \frac{1}{2} N_f.$$

In [4] it was proved that for any $f \in \mathcal{F}_n$ the absolute value of S_f is at most $2^{3n/2}$ with equality if and only if f is self-dual $\left(+2^{3n/2}\right)$ and anti-self-dual $\left(-2^{3n/2}\right)$ bent function. That is the maximum (minimum) value of the Rayleigh quotient of a Boolean function in an even number of variables is attainable on self-dual (anti-self-dual) bent functions and only them, thus providing a criterion for (anti-)self-duality in terms of the Rayleigh quotient values.

In article [8] the operations on Boolean functions that preserve bentness and the Rayleigh quotient were given. Namely, it was proved that for any $f \in \mathcal{B}_n, L \in \mathcal{O}_n, c \in \mathbb{F}_2^n, d \in \mathbb{F}_2$ the functions $g, h \in \mathcal{B}_n$ defined as $g(x) = f(Lx) \oplus d$ and $h(x) = f(x \oplus c) \oplus \langle c, x \rangle$ provide $N_g = N_f$ and $N_h = (-1)^{\langle c,c \rangle} N_f$.

The mentioned operations are isometric mappings from $\mathcal{I}_n$. The complete characterization of isometric mappings that preserve the Rayleigh quotient as well as change it was given in [19].

Theorem 20. *If $n \geqslant 4$ then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ preserves the Rayleigh quotient if and only if it preserves self-duality.*

Theorem 21. *If $n \geqslant 4$ then isometric mapping $\varphi_{\pi,g} \in \mathcal{I}_n$ changes the sign of the Rayleigh quotient if and only if it is a bijection between $\mathrm{SB}^+(n)$ and $\mathrm{SB}^-(n)$.*

7 Conclusion

In this work we considered metrical properties of the set of bent functions and its subset of functions which coincide with their duals. The group of automorphisms and metrical complements of these sets are described. We also gave some general metrical properties of the set of self-dual bent functions and considered an iterative construction.

An interesting question is the characterization of isometric mappings preserving bentness and self-duality, which are not necessarily automorphisms of the set of all Boolean functions.

References

[1] Canteaut A., Charpin P., "Decomposing bent functions", *IEEE Trans. Inform. Theory*, **49**:8 (2003), 2004–2019.

[2] Canteaut A., Daum M., Dobertin H., Leander G., "Finding nonnormal bent functions", *Discrete Appl. Math.*, **154**:2 (2006), 202–218.

[3] Carlet C., "Boolean functions for cryptography and error correcting code", *In: Crama Y., Hammer P.L. (eds.) Boolean Models and Methods in Mathematics, Computer Science, and Engineering. Cambridge University Press, Cambridge*, 2010, 257–397.

[4] Carlet C., Danielson L.E., Parker M.G., Solé. P., "Self-dual bent functions", *Int. J. Inform. Coding Theory*, **1** (2010), 384–399.

[5] Carlet C., Mesnager S., "Four decades of research on bent functions", *Des. Codes Cryptogr.*, **78**:1 (2016), 5–50.

[6] Climent J.-J., Garcia F.J., Requena V., "A construction of bent functions of $n+2$ variables from a bent function of n variables and its cyclic shifts", *Algebra*, **2014** (2014).

[7] Cusick T.W., Stănică P., *Cryptographic Boolean functions and applications*, Acad. Press, London, 2017, 288.

[8] Danielsen L.E., Parker M.G., Solé P., "The Rayleigh quotient of bent functions", *Springer Lect. Notes in Comp. Sci.*, **5921** (2009), 418–432.

[9] Dempwolff U., "Automorphisms and Equivalence of Bent Functions and of Difference Sets in Elementary Abelian 2-Groups", *Commun. Algebra*, **34**:3 (2006), 1077–1131.

[10] Dillon J., "Elementary Hadamard Difference Sets", 1974, PhD. dissertation.

[11] Feulner T., Sok L., Solé P., Wassermann A., "Towards the Classification of Self-Dual Bent Functions in Eight Variables", *Des. Codes Cryptogr.*, **68**:1 (2013), 395–406.

[12] Janusz G.J., "Parametrization of self-dual codes by orthogonal matrices", *Finite Fields Appl.*, **13**:3 (2007), 450–491.

[13] Hou X.-D., "New Constructions of Bent Functions", *Journal of Combinatorics, Information and System Sciences*, International Conference on Combinatorics, Information Theory and Statistics, **25**, 2000, 173–189.

[14] Hou X.-D., "Classification of self dual quadratic bent functions", *Des. Codes Cryptogr.*, **63**:2 (2012), 183–198.

[15] Hyun J.Y., Lee H., Lee Y., "MacWilliams duality and Gleason-type theorem on self-dual bent functions", *Des. Codes Cryptogr.*, **63**:3 (2012), 295–304.

[16] Kolomeec N.,, "The Graph of Minimal Distances of Bent Functions and Its Properties", *Des. Codes Cryptogr.*, **85**:3 (2017), 1–16.

[17] Kutsenko A.V., "The Hamming Distance Spectrum Between Self-Dual Maiorana-McFarland Bent Functions", *Journal of Applied and Industrial Mathematics*, **12**:1 (2018), 112–125.

[18] Kutsenko A., "Metrical properties of self-dual bent functions", *Des. Codes Cryptogr.*, **88**:1 (2020), 201–222.

[19] Kutsenko A., "The group of automorphisms of the set of self-dual bent functions", *Cryptogr. Commun.*, 2020, accepted, DOI: 10.1007/s12095-020-00438-y (see preprint on https://eprint.iacr.org/2019/1408).

[20] Luo G., Cao X., Mesnager S., "Several new classes of self-dual bent functions derived from involutions", *Cryptogr. Commun.*, **11**:6 (2019).

[21] MacWilliams F. J., Sloane N. J. A., "The Theory of Error-Correcting Codes", *Amsterdam, New York, Oxford: North-Holland*, 1983, 782.

[22] Markov A. A., "On transformations without error propagation", *Selected Works, Vol. II: Theory of Algorithms and Constructive Mathematics. Mathematical Logic. Informatics and Related Topics, MTsNMO, Moscow*, 2003, 70–93, In Russian.

[23] McFarland R. L., "A family of difference sets in non-cyclic groups", *J. Combin. Theory Ser. A*, **15**:1 (1973), 1–10.

[24] Mesnager S., "Several New Infinite Families of Bent Functions and Their Duals", *IEEE Trans. Inf. Theory*, **60**:7 (2014), 4397–4407.

[25] Mesnager S., *Bent Functions: Fundamentals and Results*,, Springer, Berlin, 2016, 544 p.

[26] Oblaukhov A., "A lower bound on the size of the largest metrically regular subset of the Boolean cube", *Cryptogr. Commun.*, **11**:4 (2019), 777–791.

[27] Preneel B., Van Leekwijck W., Van Linden L., Govaerts R., Vandewalle J., "Propagation characteristics of Boolean functions", *Advances in Cryptology-EUROCRYPT, Lecture Notes in Computer Science*, **473**, Springer, Berlin, Heidelberg, 1990, 161–173.

[28] Rothaus O.S., "On bent functions", *J. Combin. Theory. Ser. A*, **20**:3 (1976), 300–305.

[29] Sok L., Shi M., Solé. P., "Classification and Construction of quaternary self-dual bent functions", *Cryptogr. Commun.*, **10**:2 (2018), 277–289.

[30] Stănică P., Sasao T., Butler J.T.,, "Distance duality on some classes of Boolean functions", *J. Combin. Math. and Combin. Computing*, **107** (2018), 181–198.

[31] Tokareva N.N.,, "The group of automorphisms of the set of bent functions", *Discrete Mathematics and Applications*, **20**:5 (2010), 655–664.

[32] Tokareva N.N.,, "On the number of bent functions from iterative constructions: lower bounds", *Adv. Math. Commun.*, **5**:4 (2011), 609–621.

[33] Tokareva N., "Duality between bent functions and affine functions", *Discrete Math.*, **312**:3 (2012), 666–670.

[34] Tokareva N.N.,, "On decomposition of a Boolean function into sum of bent functions", *Siberian Electronic Mathematical Reports*, **11** (2014), 745–751.

[35] Tokareva N.N.,, "On Decomposition of a Dual Bent Function into Sum of Two Bent Functions", *Prikl. Diskretn. Mat.*, **26**:4 (2014), 59–61, In Russian.

[36] Tokareva N., *Bent Functions, Results and Applications to Cryptography*, Acad. Press. Elsevier, 2015, 230 p.

The general universal model of blockchain technology based on an analysis of some implementations

Polina Sazonova
Novosibirsk State University, JetBrains Research Laboratory
Sobolev Institute of Mathematics
Novosibirsk, Russia
Email: p.sazonova@nsu.ru

***Abstract*—First imlementation of blockchain technology was appeared in 2008, and 12 years later more than 2000 different implementations of it have appeared. After deep analysis we found that approaches for development blockchain technologies is fragmented, there are no common system of concepts and general model of technology. In this article we want to propose the general universal model and system of concepts for the blockchain technology irrespective of differenced of some implementations. Our approach is based on a technical analysis of the popular blockchains. The results of this work can be used by architects of new blockchains implementstions, by researchers to achieve theirs goals and also in educational process.**

I. Introduction

A. Blockchain definition

BLOCKCHAIN technology has become popular due to the its properties such as openness, immutability, inability to delete stored data, decentralization and the ability to make decisions in an untrusted environment between equal participants in this network without the participation of a trusted party (trusted centre). Thus, blockchain uses in a wide variety of subject areas, especially in logistics, banking and public administration.

Blockchain is a type of decentralized system that collects, stores and manages data, in which:

- consensus will be reached in an untrusted environment;
- transactions are stored in a data structure called blocks, and each subsequent block stores the value of the hash function from the contents of the previous one;
- copies of the blockchain are stored at the same time by all its users and are automatically updated.

In this work, under the blockchain is meant a system that uses a chain of blocks as a technology for storing data. It provides ensures the immutability and integrity of the data stored in the blocks. Unlike centralized systems, where consensus can be achieved through a central node, blockchain technology allows to reach consensus in decentralized environment. Moreover, in the blockchain system, consensus can be reached when the network nodes are not authorized. It means that the probability of malicious nodes or Byzantine nodes [1] appearing on the network is increase. In decentralized networks with unauthorized (untrusted) nodes, a Sybil [2] attack may occur. It can happens when the node performing the calculations connects only to nodes controlled by the attacker, which entails incorrect behavior and consensus in making a decision that is beneficial to the attacker. Blockchain technology allows to make the right decisions in a decentralized network with untrusted nodes, provided that 51% of the nodes are not intruders.

This work was supported by Math Centre in Academgorodok by agreement of The Ministry of Science and Higher Education of the Russian Federation number 075-15-2019-1613 and by JetBrains Research Cryptography Laboratory.

B. Introduction to history

The first practical implementation of blockchain technology was done in 2008, it was described in the article by S. Nakomoto about digital monetary system Bitcoin [3]. Bitcoin is a protocol for exchanging digital money in a decentralized untrusted environment that allows to make transactions without the participation of third parties (trusted centre).

But before the publication of this article, it was made lots of reseaches influented over on the blockchain technology appearing. In 1982 D. Chaum proposed the blind signature algorithm and introduced the concept of digital money [4]. S. Haber and S. Shtornetta presented a theoretical description of the system for certifying immutability of documents, built on timestamps in 1991 [5]. The Proof of Work (PoW) mechanism was proposed by A. Back in the Hashcash project to prevent [6] spamming. The idea of smart contracts was proposed by N. Szabo in 1996 [7]. N. Szabo also proposed a protocol for digital money Bit-gold in 1998, which was published in 2005 [8]; it was based on bit-chain computation and used the PoW consensus mechanism. But the system was not implemented in practice and was vulnerable to the Sybil attack.

However, the first implementation of blockchain technology was created only as a part of the Bitcoin cryptocurrency project. Subsequently, new cryptocurrency systems began to appear, similar to Bitcoin. It was added data hiding mechanisms, such as in Zcash [9], transaction acceleration mechanisms, such as in Litecoin [10]. Currencies were created for various purposes, for example, providing a set of alternative

DNS servers as in Namecoin [11]. The first implemented blockchain which was a platform for creating a smart contracts was Ethereum, created by V. Buterin in 2013 [12].

II. Motivation of Creation a Blockchain Technology Model

A. Statement of the Problem

An analysis of several hundred articles in Scopus on the topic of blockchain technologies showed that there are practically no scientific works that describes blockchain technology in general focused on its technical construction, covering all components of technology, regardless of specific implementations. In this direction it is worth highlighting this work [13], an overview of the blockchain technology components from the developers of the "Roadmap for the development of Distributed Ledger Technology (DLT)" in Russian Federation [14], an activity of the Geneva Telecommunication Standardization Sector Assembly (ITU) [15] and an activity of ISO/TC 307 committees [16]. But the results of most researchers work are not yet publicly available or have obvious flaws. This confirms the assumption that knowledge about technology is fragmented and the overall picture is not visible to researchers. This slows down the development of new technology implementations and makes it difficult to analyze new blockchains when we need to find real innovations, in contrast to the result of applying marketing tools.

B. Methods, Purpose and Criteria of the Developed Model

In this article the task of constructing a general universal model was to propose a model that would meet the following criteria: it would make it possible to make a universal description of current blockchain systems, answer questions about the structure of the system, and pose new questions to researchers and industry engineers. To build the model, an experimental-analytical approach was used: based on existing software implementations of the blockchain technology, the components of the technology were analyzed, then the obtained components were generalized, and a system of concepts was formulated for them. Then it was shown that each specific technology implementation corresponded to the proposed model.

To make a general universal model, five popular blockchains were analyzed, which are independent implementations of platforms for developing decentralized applications and cryptocurrencies. Among them: Bitcoin [17], Ethereum [18], NEO [19], DASH [20], EOS [21]. The choice of these technologies is due to their relevance as platforms for the development of decentralized applications, the high level of readiness of the technology for application, its developed by community to support them and the availability of satisfactory documentation. The characteristics of the selected blockchains are presented in the table (cf. table I).

To solve this problem the general model of blockchain technology was developed. This model does not depend on specific implementations. A key components of blockchain technology were defined and their definitions were supposed with the aim of eliminating disagreements of interpretations.

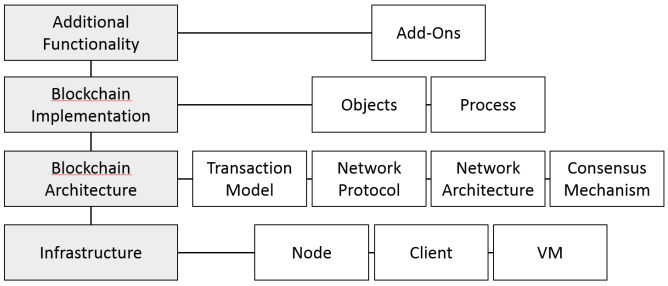


Fig. 1. Proposed blockchain technology model

The developed general model of blockchain technology is presented in the next section.

III. Proposed Blockchain Technology Model

For the five selected blockchains some documents as s technical documentations, technical concepts, «yellow papers» were analyzed. Common components that uniquely determine the blockchain technology were identified. These components are shown in the figure 1 and described in the text below.

At the **first**, basic, level of the model are the infrastructure components that ensure the functioning of the system. This is *node* – a single computer that performs actions on the network; *client* – software that implements the protocol of interaction with the blockchain; and *virtual machine (VM)* – a software system that emulates distributed work of a decentralized blockchain platform and executing decentralized applications and smart contracts.

At the **second** level, components are placed that ensure the functioning of the blockchain network. Depending on how this level is built, implementation features are established.

Network architecture – a combination of network nodes and a set of rules which uses for the the transmission of messages over the network. Blockchain networks can be single-layer or two-layer, public or private; they can have separation of nodes by roles.

Consensus Mechanism is a protocol that allows to reach an agreement between equal participants in a decentralized network. There are many implementations, but the most popular consensus is PoW, PoS, BFT and etc.

Transaction Model is a set of algorithms and features of design of the blockchain implementations that determine the method of conducting transactions and fixing the state of a distributed system. Currently, there are only two models uses in blockchains - UTXO or account model.

Network Protocol - the rules which uses for transmitted data over the network.

At the **third** level, objects and processes are located. This level arrangement depends on the implementation of the previous level. To begin with, we list **objects**, the presence of which is uniquely determined the blockchain technology.

TABLE I
CHARACTERISTICS OF THE INVESTIGATED BLOCKCHAINS

Blockchain	Transaction validation speed	Block size	One block creation speed	Bandwidth
Bitcoin	78 Min.	1 Mb	10 Min.	3 TPS
Ethereum	6 Min.	1 Mb	15 Sec.	20 (PoW), 400 (PoA) TPS
EOS	1,5 Sec.	About 1 Mb	1 Sec.	50000 TPS
NEO	15 Sec.	About 1 Mb	15 Sec.	1000-10000 TPS
DASH	15 Min.	2 Mb	1 Sec.	28-56 TPS

Block is a data structure uses to store data on the blockchain. The block stores transactions, network status, smart contracts, permissions to access data and other information.

Block chain is a data structure constructed by sequentially combining blocks into a chain. By storing the value of the hash function from the previous block, all blocks are strictly sequential, numbered by continuous numbering, the child block always refers to only one parent block.

Transaction is the minimum logically meaningful operation of the transfer or exchange of assets that makes sense and can only be completed in full. A transaction can transfer messages, actions, create a contract, and more.

Address (account, account) is a structure for identifying an active object on the network. Addresses uniquely determine the sender and recipient of the assets transferred to the blockchain network, all actions of the user in the network are associated with the address. Depending on the blockchain, the address can be either a string or a data structure, it can be associated with a user or with a smart contact.

Smart contract is a set of formalized rules implemented in the form of program code, the execution of which entails some events in the real world or digital systems. Smart contracts are not a mandatory component of the blockchain network, however, as practice has shown, contracts have become the main functional element of blockchain technology. Depending on the structure of the blockchain, smart contracts can be implemented either in Turing-complete languages or non-Turing-complete ones.

The objects listed above are part of the processes. The main **processes** taking place in the blockchain network are presented below.

Transactional life cycle: transaction signing process; broadcasting over the network; transaction verification; transaction completion. *Including a transaction in a block*: process of taking a set of transactions for a block; transaction validation; block signing process; sending a block to the network; block fixing in a common chain. *Network Maintenance*: consensus mechanism; network complexity regulation; selecting a chain that continues the block of several branches; payment for computing resources.

The **fourth** level defines additional functionality for blockchain networks that do not affect the internal architecture of the technology, but significantly expand its functionality. For example, mechanisms that provide increased speed and confidentiality of transactions, mechanisms for off-chain transactions, modules that protect blockchain against attacks by quantum computers, and others.

IV. CONCLUSIONS

After analyzing the blockchain implementations and building model as a result, we can offer *a method for considering each new technology being developed*. To analyze the new blockchain implementation, first of all, we should pay attention to the transaction model. Currently, only two models are presented - UTXO and the accounts model. The transaction model affects on: the structure of blockchain blocks, the structure of addresses (accounts), the existence of smart contracts in this blockchain and the principles of their construction, approaches to fixing the state of the system. Next, we should pay attention to the number of layers in the blockchain network, identify the purpose of each of the layers, consider the consensus mechanisms used in each layer. This information will give us an understanding of the transaction validation process – we can assume the bounds of transaction confirmation rate and network bandwidth. Based on this, we can suppose the requirenments to the necessary infrastructure to provide the network. The transactions rate is determined by the consensus mechanism, by the number of nodes involved in the transaction validation process and by the principles of working with orphaned blocks. The more stronger requirenments to network decentralization, the lower the transactions speed. The ability to create smart contracts is determined by the transaction model.

Using the results of this research we can *explain approaches to the implementation of specific blockchain technologies*. After researches we suppose that the majority of blockchain implementations are based on Bitcoin and Ethereum construction, and subsequently they were supplemented by some improvements at different levels. According to data obtained from open sources, it seems that the NEO blockchain consist of configuration of networks based on UTXO models and account models. We suppose that it makes in order to smooth out the limitations of the Bitcoin network, taken as the basis for NEO blockchain. This assumption was also made because the duplicate assets CNEO and CGAS seems artificial in these network. There is an assumption that the EOS and NEO blockchains are not blockchains, since the blockchain operates in an untrusted environment by definition, but for these networks the main transaction validators are authorized nodes, which suggests the centralization of these networks.

The Dash blockchain ensures data confidentiality and transaction speed through mechanisms operating at the fourth level of the blockchain model.

V. Results

As a result of this work, the general model of blockchain technology was proposed. This model allows to make a universal description of current blockchains, answer some questions about components and links between it in the system, and pose new questions to researchers. In this work, it was proved that the proposed model does not depend on specific implementations of the five selected blockchains and suggest methods for considering each new blockchain implementation and explain approaches to the implementation. In the future, it is planned to investigate a larger number of different blockchains in order to confirm the correctness of the model and its quality, also we plan to show connections of blockchain technology to the environment.

References

[1] L. Lamport, M. Pease, R. Shosta. "The Byzantine Generals Problem." *ACM Transactions on Programming Languages and Systems 4*, p. 3, pp. 382-401, 1982.

[2] J. R. Douceur. "The sybil attack." *International workshop on peer-to-peer systems*. Springer, Berlin, Heidelberg, 2002, pp. 251-260.

[3] S. Nakamoto. "Bitcoin: A Peer-to-Peer Electronic Cash System." The Cryptography Mailing List, 2008, https://bitcoin.org/bitcoin.pdf.

[4] D. Chaum. "Blind Signatures for Untraceable Payments." *Advances in Cryptology Proceedings of Crypto 82*, Plenum, 1982, pp. 199-203, 1982.

[5] S.Haber, W.S. Stornetta. "How to time-stamp a digital document." *J. Cryptology 3*. 1991, pp. 99–111.

[6] A. Back. Mail "Hash cash postage implementation". The Cypherpunks Mailing List. https://cypherpunks.venona.com/date/1997/03/msg00774.html.

[7] N. Szabo. "Smart Contracts: Building Blocks for Digital Markets." A partial rewrite of the article which appeared in Extropy No 16, 1996, http://www.alamut.com/subj/economics/nick_szabo/smartContracts.html.

[8] N. Szabo. "Bit gold." Unenumerated: N. Szabo's blog, 2005, https://web.archive.org/web/20060329122942/http://unenumerated.blogspot.com/2005/12/bit-gold.html.

[9] Zcash - a privacy-protecting, digital currency, https://z.cash/.

[10] Litecoin - decentralised money, https://litecoin.com/en/.

[11] Namecoin - a trust anchor for the Internet, https://www.namecoin.org/.

[12] Ethereum - a global, open-source platform for decentralized applications, https://ethereum.org/ru/.

[13] H.Y. Paik, X. Xu, H. M. N. Dilum Bandara, S. U. Lee, S. K. Lo. "Analysis of Data Management in Blockchain-Based Systems: From Arcitecture to Governance." *IEEE Access*, 2019, t.7, pp. 186091-186107, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8938787.

[14] Ministry of Digital Development, Communications and Mass Media of the Russian Federation. "Roadmap for the development of Distributed Ledger Technology (DLT)." https://digital.gov.ru/ru/documents/6670/.

[15] ITU's Telecommunication Standardization Sector (ITU-T). "ITU-T Focus Group on Application of Distributed Ledger Technology." https://www.itu.int/en/ITU-T/focusgroups/dlt/Pages/default.aspx#.

[16] Technical cometees ISO/TC 307. "Blockchain and distributed ledger technologies." https://www.iso.org/committee/6266604.html.

[17] Bitcoin developers documentation, https://developer.bitcoin.org/.

[18] G. Wood. Ethereum: a secure decentralized generalized transaction ledger, https://ethereum.github.io/yellowpaper/paper.pdf.

[19] Technical Specification for NEO Blockchain, https://github.com/neoresearch/yellowpaper.

[20] E. Duffield, D. Diaz. Dash: A Payments-Focused Cryptocurrency. https://github.com/dashpay/dash/wiki/Whitepaper.

[21] EOS Developer Portal, https://developers.eos.io/.

On relationship between quaternary and Boolean bent functions

Shaporenko Alexander
Sobolev Institute of Mathematics
Novosibirsk State University
Laboratory of Cryptography JetBrains Research
Novosibirsk, Russia
shaporenko.alexandr@gmail.com

***Abstract*—In this paper, the relationship between quaternary and Boolean bent functions is studied. The importance of Boolean bent functions in symmetric cryptography stems from linear cryptanalysis of stream ciphers. In that context bent functions are the ones which are the worst approximated by affine functions. There are also connections between bent functions and distinct objects of coding theory such as Reed-Muller and Kerdock codes. The concept of bent functions was generalized for q-ary functions $g : \mathbb{Z}_q^n \to \mathbb{Z}_q$. Function $g : \mathbb{Z}_4^n \to \mathbb{Z}_4$ is called a quaternary function in n variables. Any quaternary function g in n variables can be uniquely represented for any $x, y \in \mathbb{Z}_2^n$ as $g(x + 2y) = a(x, y) + 2b(x, y)$ with a and b be Boolean funtions in $2n$ variables. A representation of the Walsh–Hadamard coefficients of a quaternary function is obtained in terms of the coefficients of Boolean functions b and $a \oplus b$. A series of statements are proved showing that the bentness property of a quaternary function g doesn't directly depend on the bentness of Boolean functions b and $a \oplus b$. The number of quaternary bent functions in one and two variables is obtained with a description of properties of Boolean functions b and $a \oplus b$. It was proved that bentness of a quaternary function implies that b and $a \oplus b$ are nonlinear. A simple construction of quaternary bent functions in any number of variables is presented.**

***Index Terms*—cryptography, quaternary functions, Boolean functions, bent functions**

I. Introduction

A Boolean function in even number of variables is called bent if it is maximal nonlinear [1]. Nonlinearity is an important property in cryptography. Ciphers in which functions with high nonlinearity are used as components are more resistant to linear cryptanalysis [2] method because bent fucntions are bad in being approximated by affine functions. Bent functions were used in design of the block cipher CAST as coordinate functions of S-blocks [3]. The nonlinear feedback polynomial of the NFSR (nonlinear feedback shift register) of the stream cipher Grain is constructed as the sum of a linear function and a bent function [4]. There are also connections between bent functions and distinct objects of coding theory such as *Reed-Muller and Kerdock codes* [5].

In [6] *q-ary* ($g : \mathbb{Z}_q^n \to \mathbb{Z}_q$) bent functions were defined for $q > 1$. The study of such functions was due to the desire of authors to summarize the results of [7] on the use of Boolean bent functions in CDMA (Code Division Multiple Access) systems. There are also some works related to extension of usual linear cryptanalysis such as $\mathbb{Z}_4$-linear cryptanalysis [8].

In [9] direct links between Boolean bent functions and another generalization of Boolean bent functions ($f : \mathbb{Z}_2^n \to \mathbb{Z}_q$) [10] were explored. We continue this work.

In this paper, the relationship between quaternary ($g : \mathbb{Z}_4^n \to \mathbb{Z}_4$) and Boolean bent functions is studied. It was proven that the bentness property of a quaternary function $g(x + 2y) = a(x, y) + 2b(x, y)$ doesn't directly depend on the bentness of Boolean functions b and $a \oplus b$. The number of quaternary bent functions in one and two variables is obtained with a description of properties of Boolean functions b and $a \oplus b$. A simple construction of quaternary bent functions in any number of variables is presented.

Let $\mathbb{Z}_2 = \{0, 1\}$ and $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Denote by $\mathbb{Z}_2^n$ the n-dimensional vector space over $\mathbb{Z}_2$ and by $\mathbb{Z}_4^n$ the n-dimensional vector space over $\mathbb{Z}_4$. Let $\langle x, y\rangle$ be an inner product of vectors with summation modulo 2 (denote by $\oplus$) and $x.y$ be an inner product of vectors over $\mathbb{Z}_4$. *The Walsh–Hadamard transform of Boolean function* f in n variables is the integer-valued function on $\mathbb{Z}_2^n$ defined as

$$W_f(x) = \sum_{y \in \mathbb{Z}_2^n} (-1)^{\langle x,y\rangle \oplus f(y)} \text{ for every } x \in \mathbb{Z}_2^n.$$

Numbers $W_f(x)$ are called *Walsh–Hadamard coefficients* of a Boolean function f. A *bent function* is a Boolean function in n variables (n is even) such that $|W_f(x)| = 2^{n/2}$ for every $x \in \mathbb{Z}_2^n$.

Let g be a function from $\mathbb{Z}_4^n$ to $\mathbb{Z}_4$. *The Walsh–Hadamard transform of a quaternary function* g is defined as follows

$$W_g(x) = \sum_{y \in \mathbb{Z}_4^n} i^{x.y+g(y)} \text{ for every } x \in \mathbb{Z}_4^n,$$

where '+' is the addition over $\mathbb{Z}_4$. A quaternary function g in n variables is called *quaternary bent function* if $|W_g(x)| = 4^{n/2}$ for every $x \in \mathbb{Z}_4^n$.

Functions $f(x) = \langle a, x\rangle$ with $a, x \in \mathbb{Z}_2^n$ are called linear Boolean functions in n variables.

The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

II. Quaternary bent functions in one and two variables

Results presented in this section were obtained via exhaustive search of all quaternary functions in one (4^4 in total) and two (4^{16} in total) variables.

Note that any quaternary function g in n variables can be uniquely represented as follows

$$g(x+2y) = a(x,y) + 2b(x,y) \text{ for every } x, y \in \mathbb{Z}_2^n.$$

Here '+' is the addition over $\mathbb{Z}_4$ and a, b are Boolean functions in $2n$ variables.

Proposition 1. *Let $g(x+2y) = a(x,y) + 2b(x,y)$ be a quaternary bent function in one variable with $x, y \in \mathbb{Z}_2$ and a, b be Boolean functions in two variables. Then b and $a \oplus b$ are bent functions. Here '+' is the addition over $\mathbb{Z}_4$.*

Proposition 2. *For every quaternary function $g(x+2y) = a(x,y) + 2b(x,y)$ in one variable with $x, y \in \mathbb{Z}_2$ it is true that g is a quaternary bent function if and only if b is a bent function and a does not depend on y, i.e. $a(x,y) = 0$, 1, x or $x \oplus 1$. Here '+' is the addition over $\mathbb{Z}_4$.*

Computer search shows that the number of quaternary bent functions in one variable is equal to 32.

There are 200704 quaternary bent functions in 2 variables. Among them there are 98304 fuctions such that none of Boolean functions a, b and $a \oplus b$ is a bent function but for 3072 of them a is a linear Boolean function. There are 36864 quaternary bent functions such that b and $a \oplus b$ are bent functions, while for 33792 of them a is a nonlinear function, and for 2304 and 768 functions a is a linear function or constant respectively. The number of quaternary bent function in 2 variables with each of a, b and $a \oplus b$ being a bent function is equal to 16384. For the remaining 49152 functions, a is a bent function and b and $a \oplus b$ are nonlinear Boolean functions.

For functions in three and more variables an exhaustive search is too hard (there are 2^{128} quaternary functions in three variables).

III. Connection between quaternary and Boolean bent functions

Results presented in this section shows that the bentness property of a quaternary function g doesn't directly depend on the bentness of Boolean functions b and $a \oplus b$. The following two lemmas are instrumental in what follows.

Lemma 1.

$$2(\bigoplus_{i=1}^{n} z_i) = 2z_1 + ... + 2z_n.$$

Here '+' is the addition over $\mathbb{Z}_4$.

Proof. Let $|\{i : z_i = 1\}| = k$. We have two cases:

1) *k is even.* Then

$$2(\bigoplus_{i=1}^{n} z_i) = 2 \cdot 0 = 0,$$
$$2z_1 + ... + 2z_n = 2 \cdot k \bmod 4 = 2 \cdot (2 \cdot l) \bmod 4 =$$
$$= 4 \cdot l \bmod 4 = 0,$$
$$l \in \mathbb{N} \cup \{0\}.$$

2) *k is odd.* Then

$$2(\bigoplus_{i=1}^{n} z_i) = 2 \cdot 1 = 2,$$
$$2z_1 + ... + 2z_n = 2 \cdot k \bmod 4 = 2 \cdot (2 \cdot l + 1) \bmod 4 =$$
$$= 4 \cdot l + 2 \bmod 4 = 2,$$
$$l \in \mathbb{N} \cup \{0\}.$$

□

Lemma 2. *There is a relation between Walsh–Hadamard coefficients of g, b and $a \oplus b$*

$$W_g(x+2y) =$$
$$= \frac{1}{2}(W_b(x \oplus y, x) + W_{a \oplus b}(y, x) - 2c_1 - 2d_1) +$$
$$+ \frac{i}{2}(W_b(y, x) - W_{a \oplus b}(x \oplus y, x) - 2c_2 + 2d_2),$$

with

$$c_1 = \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle \oplus \langle x,x' \rangle}$$
$$c_2 = \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle}$$
$$d_1 = \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus a(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle}$$
$$d_2 = \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus a(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle \oplus \langle x,x' \rangle}.$$

Here $X = \{x' | \langle x, x' \rangle = x.x'\}$.

Proof. In what follows, '+' in exponent denotes addition over $\mathbb{Z}_4$.

For function $g(x+2y) = a(x,y) + 2b(x,y)$ the Walsh–Hadamard coefficient is represented as follows

$$W_g(x+2y) = \sum_{x' \in \mathbb{Z}_2^n, y' \in \mathbb{Z}_2^n} i^{(x+2y).(x'+2y') + a(x',y') + 2b(x',y')}.$$

From Lemma 1 we get that $2\langle x'', x''' \rangle = 2x''.x'''$ for every $x'', x''' \in \mathbb{Z}_2^n$. Then

$$(x+2y).(x'+2y') =$$

$$= \begin{cases} \langle x, x' \rangle + 2\langle x, y' \rangle + 2\langle y, x' \rangle, & \text{if } x.x' = \langle x, x' \rangle, \\ \langle x, x' \rangle + 2\langle x, y' \rangle + 2\langle y, x' \rangle + 2, & \text{if } x.x' \neq \langle x, x' \rangle. \end{cases}$$

Let $X = \{x' | x.x' = \langle x, x' \rangle\}$ then

$$\begin{aligned}
&W_g(x+2y) = \\
&= \sum_{x' \in X, y' \in \mathbb{Z}_2^n} i^{\langle x,x' \rangle + 2\langle x,y' \rangle + 2\langle y,x' \rangle + a(x',y') + 2b(x',y')} \\
&+ \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} i^{\langle x,x' \rangle + 2\langle x,y' \rangle + 2\langle y,x' \rangle + a(x',y') + 2b(x',y') + 2} \\
&= \sum_{x \in X, y' \in \mathbb{Z}_2^n} (-1)^{\langle x,y' \rangle + \langle y,x' \rangle + b(x',y')} i^{\langle x,x' \rangle + a(x',y')} \\
&- \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{\langle x,y' \rangle + \langle y,x' \rangle + b(x',y')} i^{\langle x,x' \rangle + a(x',y')}.
\end{aligned}$$

Here we use the standard maps $\beta, \gamma : \mathbb{Z}_4 \to \mathbb{Z}_2$ defined as

$$\begin{aligned}
&\beta : 0, 1 \to 0 \text{ and } \beta : 2, 3 \to 1; \\
&\gamma : 0, 2 \to 0 \text{ and } \gamma : 1, 3 \to 1.
\end{aligned}$$

For any $t \in \mathbb{Z}_4$ it holds

$$i^t = (-1)^{\beta(t)} \left(\frac{1 + (-1)^{\gamma(t)}}{2} + \frac{1 - (-1)^{\gamma(t)}}{2} i \right).$$

Using this formula for $t = \langle x, x' \rangle + a(x', y')$ and the fact that $\gamma(\langle x, x' \rangle + a(x', y')) = \langle x, x' \rangle \oplus a(x', y')$ we get that

$$\begin{aligned}
&W_g(x+2y) = \\
&= \frac{1}{2}(S_1 + S_2 - S_3 - S_4) + \frac{i}{2}(S_1 - S_2 - S_3 + S_4),
\end{aligned}$$

with

$$\begin{aligned}
S_1 &= \\
&\sum_{x' \in X, y' \in \mathbb{Z}_2^n} (-1)^{z + \beta(\langle x,x' \rangle + a(x',y'))} \\
S_2 &= \\
&\sum_{x' \in X, y' \in \mathbb{Z}_2^n} (-1)^{z + \langle x,x' \rangle + a(x',y') + \beta(\langle x,x' \rangle + a(x',y'))} \\
S_3 &= \\
&\sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{z + \beta(\langle x,x' \rangle + a(x',y'))} \\
S_4 &= \\
&\sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{z + \langle x,x' \rangle + a(x',y') + \beta(\langle x,x' \rangle + a(x',y'))}.
\end{aligned}$$

Here $z = \langle x, y' \rangle + \langle y, x' \rangle + b(x', y')$.

Let $M_\delta = \{x' | \langle x.x' \rangle = \delta\}$ for $\delta \in \mathbb{Z}_2$. Therefore, we get $\mathbb{Z}_2^n = (M_0 \cap X) \cup (M_1 \cap X) \cup (M_0 \cap \mathbb{Z}_2^n \setminus X) \cup (M_1 \cap \mathbb{Z}_2^n \setminus X)$. Let us divide every sum S_1, S_2, S_3 and S_4 into two sums $\sum_{x' \in M_0, y' \in \mathbb{Z}_2^n}$ and $\sum_{x' \in M_1, y' \in \mathbb{Z}_2^n}$.

Note that $\beta(a(x', y') + \langle x, x' \rangle)$ is equal to 0 or $a(x', y')$ for $x' \in M_0$ and $x' \in M_1$ respectively. Thus, after grouping items we obtain

$$\begin{aligned}
&S_1 + S_2 - S_3 - S_4 = \\
&= \sum_{x' \in X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle \oplus \langle x,x' \rangle} + \\
&+ \sum_{x' \in X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus a(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle} - \\
&- \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle \oplus \langle x,x' \rangle} - \\
&- \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus a(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle}.
\end{aligned}$$

Then

$$\begin{aligned}
&S_1 - S_2 - S_3 + S_4 = \\
&= \sum_{x' \in X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle} - \\
&- \sum_{x' \in X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus a(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle \oplus \langle x,x' \rangle} - \\
&- \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle} + \\
&+ \sum_{x' \in \mathbb{Z}_2^n \setminus X, y' \in \mathbb{Z}_2^n} (-1)^{b(x',y') \oplus a(x',y') \oplus \langle x,y' \rangle \oplus \langle y,x' \rangle \oplus \langle x,x' \rangle}.
\end{aligned}$$

The reason why addition over $\mathbb{Z}_4$ in exponent was replaced by addition over $\mathbb{Z}_2$ is because $(-1)^{z \bmod 4} = (-1)^{z \bmod 2}$ for any $z \in \mathbb{Z}$.

Adding and subtracting c_1 and d_1 in the first equation and c_2 and d_2 in the second one, we obtain the necessary. □

Theorem 1. *Let $g(x + 2y) = a(x, y) + 2b(x, y)$ be a quaternary bent function with $x, y \in \mathbb{Z}_2^n$ and a, b be Boolean functions in $2n$ variables. Then b and $a \oplus b$ are nonlinear functions for any $n \geqslant 1$.*

Proof. There are three possible values of Walsh–Hadamard coefficients of a linear Boolean function in $2n$ variables, $\{0, \pm 2^{2n}\}$.

From Lemma 2 we get that

$$\begin{aligned}
&W_g(2y) = \\
&= \frac{1}{2}(W_b(y, 0) + W_{a \oplus b}(y, 0)) + \frac{i}{2}(W_b(y, 0) - W_{a \oplus b}(y, 0)),
\end{aligned}$$

with $y \in \mathbb{Z}_2^n$. Note that the reason why there are no coefficients c_1, c_2, d_1 and d_2 is because the set $\mathbb{Z}_2^n \setminus X$ is empty for $x = \{0, ..., 0\}$.

For any complex number $z = x + iy$ it is known that $|z|^2 = x^2 + y^2$. We know that $|W_g(2y)| = 4^{n/2}$ because g is a quaternary bent. Hence,

$$(W_b(y, 0) + W_{a \oplus b}(y, 0))^2 + (W_b(y, 0) - W_{a \oplus b}(y, 0) = 4 \cdot 4^n.$$

From [6] we know that each of Walsh–Hadamard coefficients of quaternary bent function can be expressed as

$$W_g(z) = 4^{n/2} i^{h(x)} \text{ for every } z \in \mathbb{Z}_4^n$$

for some quaternary function h. It means that there is only real or imaginary part of $W_g(2y)$. Thus, we get that there are two possible cases

$$\begin{cases}(W_b(y,0)+W_{a\oplus b}(y,0))^2=0\\(W_b(y,0)-W_{a\oplus b}(y,0))^2=4\cdot 4^n.\end{cases}$$

or

$$\begin{cases}(W_b(y,0)+W_{a\oplus b}(y,0))^2=4\cdot 4^n\\(W_b(y,0)-W_{a\oplus b}(y,0))^2=0.\end{cases}$$

From the first system we get

$$\begin{cases}W_b(y,0)=-W_{a\oplus b}(y,0),\\(2\cdot W_b(y,0))^2=4\cdot W_b(y,0)^2=4\cdot 4^n.\end{cases}$$

Hence,

$$W_b(y,0)=-W_{a\oplus b}(y,0)=\pm 2^n.$$

One can get that by solving the second system you get

$$W_b(y,0)=W_{a\oplus b}(y,0)=\pm 2^n.$$

Therefore, $b, a\oplus b$ are nonlinear functions. □

Proposition 3. *For every $n\geqslant 1$ there exists a quaternary function $g(x+2y)=a(x,y)+2b(x,y)$ in n variables which is not bent, while b and $a\oplus b$ are Boolean bent functions in $2n$ variables.*

Proof. In what follows, '+' denotes addition over $\mathbb{Z}_4$ excepting summation of indices.

Any quaternary function g in n variables can be uniquely represented as follows

$$g(x_1+2x_{n+1},...,x_n+2x_{2n})=$$
$$=a(x_1,...,x_{2n})+2b(x_1,...,x_{2n}).$$

Let

$$b(x_1,..,x_{2n})=\bigoplus_{i=1}^{n}x_ix_{i+n},$$
$$a(x_1,..,x_{2n})=x_{n+1}\oplus c,$$

with $c\in\mathbb{Z}_2$.

From Lemma 1 we know that

$$2b(x_1,..,x_{2n})=2x_1x_{n+1}+...+2x_nx_{2n}.$$

One can see that g can be divided into sum of n quaternary functions in one variable

$$g(x_1+2x_{n+1},...,x_n+2x_{2n})=$$
$$=g_1(x_1+2x_{n+1})+...+g_n(x_n+2x_{2n}),$$
$$g_i(x_i+2x_{n+i})=a_i(x_i,x_{n+i})+2b_i(x_i,x_{n+i}),$$
$$b_i(x_i,x_{n+i})=x_ix_{n+i},$$
$$a_i(x_i,x_{n+i})=\begin{cases}x_{n+i}\oplus c, i=1,\\0,\ \text{otherwise.}\end{cases}$$

It is known that g is a quaternary bent function if and only if all of g_i are quaternary bent functions [11], $i=1,...,n$. From Proposition 2 and by the choice of a and b we get that g_1 is not quaternary bent. This completes the proof. □

Proposition 4. *For every $n\geqslant 2$ there exists a quaternary bent function $g(x+2y)=a(x,y)+2b(x,y)$ in n variables, with b and $a\oplus b$ being not bent in $2n$ variables.*

Proof. In what follows, '+' denotes addition over $\mathbb{Z}_4$ excepting summation of indices.

Any quaternary function g in n variables can be uniquely represented as follows

$$g(x_1+2x_{n+1},...,x_n+2x_{2n})=$$
$$=a(x_1,...,x_{2n})+2b(x_1,...,x_{2n}).$$

Let

$$b(x_1,..,x_{2n})=\bigoplus_{i=3}^{n}x_ix_{i+n}\oplus x_1x_{n+2}\oplus$$
$$\oplus x_2x_{n+1}\oplus x_1x_2x_{n+1},$$
$$a(x_1,..,x_{2n})=x_1x_{n+1}.$$

One can see that b can be divided into sum of $n-2$ Boolean functions in two variables and one Boolean function in four variables

$$b(x_1,...,x_{2n})=b_1(x_1,x_2,x_{n+1},x_{n+2})\oplus$$
$$\oplus b_2(x_3,x_{n+3})\oplus...\oplus b_{n-1}(x_n,x_{2n}),$$
$$b_1(x_1,x_2,x_{n+1},x_{n+2})=x_1x_{n+2}\oplus x_2x_{n+1}\oplus x_1x_2x_{n+1},$$
$$b_i(x_{i+1},x_{n+i+1})=x_{i+1}x_{n+i+1},$$
$$i=2,...,n-1.$$

It is known that b is a bent function if and only if all of b_i are bent functions [5]. Function b_1 in four variables is not bent because its degree is equal to three [5].

From Lemma 1 we know that

$$2b(x_1,..,x_{2n})=2x_3x_{n+3}+...+2x_nx_{2n}+$$
$$+2x_1x_{n+2}+2x_2x_{n+1}+2x_1x_2x_{n+1}.$$

Moreover, g can be divided into sum of $n-2$ quaternary functions in one variable and one quaternary function in two variables

$$g(x_1+2x_{n+1},...,x_n+2x_{2n})=g_1(x_1+2x_{n+1},x_2+2x_{n+2})+$$
$$+g_2(x_3+2x_{n+3})+...+g_{n-1}(x_n+2x_{2n}),$$
$$g_1(x_1+2x_{n+1},x_2+2x_{n+2})=x_1x_{n+1}+$$
$$+2x_1x_{n+2}+2x_2x_{n+1}+2x_1x_2x_{n+1},$$
$$g_i(x_{i+1}+2x_{n+i+1})=2x_{i+1}x_{n+i+1},$$
$$i=2,...,n-1.$$

From Proposition 2 functions g_i are quaternary bent functions, $i=2,...,n-1$. It was checked that the quaternary function g_1 is also bent.

It is true that g is a quaternary bent function if and only if all of g_i are bent quaternary bent functions [11], $i=1,...,n-1$. This completes the proof. □

Next hypothesis is based on the fact that the following statement is true for quaternary functions in one and two variables.

Hypothesis 1. *For a quaternary bent function* $g(x+2y) = a(x,y) + 2b(x,y)$ *in any number of variables it is true that*

$$b \text{ is a bent function } \iff a \oplus b \text{ is a bent function.}$$

IV. Construction of quaternary bent functions

Following result can be used as a simple construction for quaternary bent function in any number of variables.

Proposition 5. *For every* n *a quaternary function*

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = ax_j + x_1x_{n+1} + ... + x_nx_{2n}$$

is a quaternary bent function with $a \in \mathbb{Z}_2$, $j \in \{1, ..., n\}$ *and* '$+$' *is addition over* $\mathbb{Z}_4$.

Proof. One can see that g can be divided into sum of n quaternary functions in one variable

$$\begin{gathered} g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = \\ = g_1(x_1 + 2x_{1+n}) + ... + g_n(x_n + 2x_{2n}), \\ g_i(x_i + 2x_{i+n}) = 2x_ix_{i+n}, \text{ for } i \neq j, \\ g_j(x_j + 2x_{j+n}) = ax_j + 2x_jx_{j+n}. \end{gathered}$$

From Proposition 2 each of g_i is a quaternary bent function in one variable, therefore, g is also a quaternary bent function [11]. □

V. Conclusion and open problems

Although the results show that there is no direct connection between quaternary and Boolean bent functions it's still might be possible to connect these notions if we will ask for additional conditions like it did in Hypothesis 1.

The next step besides proving or refuting Hypothesis 1 can be generalization of all results for arbitrary q.

Author wish to thank Natalia Tokareva and Aleksandr Kutsenko for helpful advices and interest to this work.

References

[1] O. S. Rothaus, "On 'bent' functions," J. Combinat. Theory A, vol. 20, no. 3, pp. 300–305, 1976.
[2] M. Matsui, "Linear Cryptanalysis Method for DES cipher," Advances in Cryptology – Eurocrypt 1993, Springer-Verlog, Berlin, pp. 386–397.
[3] C. Adams, "Constructing symmetric ciphers using the CAST design procedure," Proc. Design, Codes, and Cryptography, vol. 12, no. 3, pp. 283–316, 1997.
[4] M. Hell, T. Johansson, A. Maximov, and W. Meier, "A stream cipher proposal: Grain-128," IEEE International Symposium on Information Theory, pp. 1614–1618, 2006.
[5] N. Tokareva, "Bent functions: results and applications to cryptography," Acad. Press. Elsevier, 2015.
[6] P. V. Kumar, R. A. Scholtz, and L. R. Welch, "Generalized bent functions and their properties," J. Combin. Theory A., vol. 40, no. 1, pp. 90–107, 1985.
[7] G. D. Olsen, R. A. Scholtz, and L. R. Welch, "Bent-function sequences," IEEE Trans. Inform. Theory, vol. 28, no. 6, pp. 858–864, 1982.
[8] M. G. Parker and H. Raddum, "Z4-Linear Cryptanalysis," NESSIE Internal Report, 27/06/2002: NES/DOC/UIB/WP5/018/1.
[9] P. Sole and N. Tokareva, "Connections between Quaternary and Binary Bent Functions," unpublished, Cryptology ePrint Archive, Report 2009/544.
[10] K-U. Schmidt, "Quaternary Constant-Amplitude Codes for Multicode CDMA," Trans. Inform. Theory, vol. 55. no. 4. pp. 1824–1832, 2009.
[11] D. Singh, M. Bhaintwal, and B. K. Singh, "Some results on q-ary bent functions," Int. J. Comput. Math., vol. 90, no. 9, pp. 1761–1773, 2013.

S-box construction based on a Boolean function and a permutation

1st Darya Zyubina
Sobolev Institute of Mathematics
Novosibirsk State University
Laboratory of Cryptography
JetBrains Research
Novosibirsk, Russia
d.zyubina@g.nsu.ru

2nd Maxim Zapolskiy
Sobolev Institute of Mathematics
Novosibirsk State University
Laboratory of Cryptography
JetBrains Research
Novosibirsk, Russia
m.zapolskii@g.nsu.ru

3rd Irina Khilchuk
Novosibirsk State University
Laboratory of Cryptography
JetBrains Research
Novosibirsk, Russia
i.khilchuk@g.nsu.ru

4th Natalia Tokareva
Sobolev Institute of Mathematics
Novosibirsk, Russia
tokareva@math.nsc.ru

***Abstract*—The paper is devoted to the construction of vectorial Boolean functions used in S-boxes. We consider the method to construct vectorial Boolean functions using Boolean functions and permutations on variables in order to simplify the construction of vectorial Boolean functions with certain cryptographic properties such as high nonlinearity, balancedness, low differential δ-uniformity and high algebraic degree. Cryptographic properties of vectorial Boolean functions constructed via our method are analyzed for the small number of variables. Necessary and sufficient conditions for bijectivity of the constructed vectorial Boolean function are determined for an arbitrary number of variables.**

***Index Terms*—Boolean function, vectorial Boolean function, S-box**

I. Introduction

S-boxes play the crucial role for providing resistance of a block cipher to different types of attacks. The major reason for this is the following: in classical and modern block ciphers the main complicated and nonlinear layer is presented namely by S-boxes. Mathematically, S-box is a vectorial Boolean function that maps n bits to m bits, or, $n \to m$. Usually, n coincides with m, moreover, it is a necessary condition for S-box to be one-to-one, i. e. bijective. Often, the number n that is considered in practice is not too big. For example, in block ciphers GOST 28147-89 [1] and Present [2] S-boxes $4 \to 4$ are used, cipher DES (ex-standard of USA) operates with S-boxes of type $6 \to 4$, modern ciphers AES [3] and GOST R 34.12-15 (known as Kuznyechik [4]) use S-boxes $8 \to 8$.

Indeed, it is very difficult to construct S-boxes of big sizes. Let us remind that the number of distinct vectorial Boolean functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ is equal to $2^{n \cdot 2^n}$. It means that even for dimension $n = 4$ there are 2^{64} distinct vectorial Boolean functions in n variables; in case $n = 6$ we can not even desire to check them all using computer search, since the number of them is 2^{384}. But in fact, we are very interested in obtaining *good* S-boxes in order to construct resistant ciphers. It is well known that some special mathematical properties of S-box, such as high nonlinearity, low differential uniformity, high algebraic immunity, etc. make a cipher with such S-box be resistant to linear, differential, algebraic and other methods of cryptanalysis. It is well known that cryptographic properties of a Boolean (vectorial) function contradict to each other, [5], [6]. That is why we seek to find vectorial Boolean functions that reach a *tradeoff* between different cryptographic properties. That is why we are obligated to use mathematical methods (and not a direct computer search) for their constructing.

In this paper we propose a simple method of constructing S-boxes using Boolean functions. We take a Boolean function f in n variables and a permutation π on n variables and construct a vectorial Boolean function F_π that maps n bits into n bits by the following rule: $F_\pi(x) = (f(x), f(\pi(x)), f(\pi^2(x)), \ldots, f(\pi^{n-1}(x)))$, where x runs through $\mathbb{F}_2^n$. Then we study the cryptographic properties of F_π with respect to f and π. In general case we give an answer when F_π is a one-to-one function; for the small number of variables we analyze what cryptographic properties of F_π we can provide. So, we invite specialists to use this construction for obtaining good S-boxes. Note that for a fixed π the number of distinct functions F_π is 2^{2^n}, i. e. is equal to the number of all Boolean functions in n variables. So, exhaustive search methods are still useful for relatively big dimensions.

The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

II. Definitions

Let $\mathbb{F}_2^n$ be the vector space of dimension n over $\mathbb{F}_2$. Let $x = (x_1, \ldots, x_n)$ be a binary vector and $\oplus$ denote the addition modulo 2 (XOR). Recall that *standard inner product* of two binary vectors is $\langle x, y \rangle = x_1 y_1 \oplus \ldots \oplus x_n y_n$.

The *Walsh — Hadamard transform* of a Boolean function f is

$$W_f(y) = \sum_{x \in \mathbb{F}_2^n} (-1)^{\langle x,y \rangle \oplus f(x)}.$$

We know that any Boolean function can be uniquely represented in its *algebraic normal form* (ANF):

$$f(x_1, \ldots, x_n) = \left(\bigoplus_{k=1}^{n} \bigoplus_{i_1, \ldots, i_k} a_{i_1, \ldots, i_k} \, x_{i_1} \cdot \ldots \cdot x_{i_k} \right) \oplus a_0,$$

where for each k indices $i_1, \ldots, i_k$ are pairwise distinct and sets $\{i_1, \ldots, i_k\}$ are exactly all different nonempty subsets of the set $\{1, \ldots, n\}$; coefficients $a_{i_1,\ldots,i_k}$, a_0 take values from $\mathbb{F}_2$. For a Boolean function f the number of variables in the longest item of its ANF is called the *algebraic degree* of a function (or briefly *degree*) and is denoted by $deg(f)$. A Boolean function is *affine*, *quadratic*, *cubic* and so on if its degree is not more than 1, or equal to 2, 3, etc.

III. Cryptographic properties of the functions

A Boolean function f in n variables is called *balanced* if it takes every value (0 or 1) the same number of times [7]. A vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is *balanced* if it takes every value of $\mathbb{F}_2^n$ equally often [6] .

The *nonlinearity* $nl(f)$ of a Boolean function f in n variables is the minimum Hamming distance between f and the set of all affine Boolean functions in n variables [7] . The *nonlinearity* $nl(F)$ of a vectorial Boolean function F is the minimal nonlinearity of all its component Boolean functions:

$$nl(F) = \min_{v \in \mathbb{F}_2^n} nl(F_v) = \min_{v \in \mathbb{F}_2^n} d(\langle v, F \rangle, \mathcal{A}_n) = \\ = \min_{v \in \mathbb{F}_2^n} \min_{g \in \mathcal{A}_n} d(\langle v, F \rangle, g),$$

where $v \neq 0$ and $\mathcal{A}_n = \{\langle a, x \rangle \oplus b : a \in \mathbb{F}_2^n, b \in \mathbb{F}_2\}$ is the class of all affine Boolean functions of n variables [6].

The *algebraic degree* of a vectorial Boolean function is the maximal algebraic degree of its component functions [6] .

Let δ be a positive even integer. A vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is *differential δ-uniform* if for every $a \neq 0$ in $\mathbb{F}_2^n$ and every $b \in \mathbb{F}_2^n$ the equation $F(x) \oplus F(x \oplus a) = b$ has not more than δ solutions [6] . The minimal possible value of δ for functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ is 2. Differential 2-uniform functions are called *APN functions*.

IV. S-box construction based on a Boolean function

Let us present the following construction of a vectorial Boolean function based on a Boolean function and a permutation. Let π be an arbitrary permutation on n elements, $\pi \in S_n$. If $x = (x_1, \ldots, x_n)$ is a binary vector then let $\pi(x)$ be a vector obtained as $\pi(x) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. Let f be a Boolean function in n variables. Define a vectorial Boolean function $F_\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ as follows

$$F_\pi(x) = (f(x), f(\pi(x)), f(\pi^2(x)), \ldots, f(\pi^{n-1}(x))).$$

In this paper we would like to study cryptographic properties of the vectorial Boolean function F_π in dependence of properties of the Boolean function f and the permutation π.

Denote by $\Delta_{\pi,n}$ the class of all vectorial Boolean functions in n variables obtained in the described manner. So,

$$\Delta_{\pi,n} = \{F_\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n : \\ F_\pi(x) = (f(x), f(\pi(x)), \ldots, f(\pi^{n-1}(x))) \\ \text{for some } f : \mathbb{F}_2^n \to \mathbb{F}_2\}.$$

Separately, we consider the special case of a permutation. Let ρ be a cyclic permutation on n coordinates, namely

$$\rho(x) = (x_n, x_1, x_2, \ldots, x_{n-1}).$$

It is a permutation that very naturally can be found in different cryptographic constructions. It reflects the principle of acting for shift registers and is used for defining rotation symmetric functions [5] applied for cryptography.

If π is identical then the class $\Delta_{\pi,n}$ we denote briefly as Δ_n. If π is a cyclic rotation, i. e. $\pi = \rho$ then we deal with class $\Delta_{\rho,n}$. Otherwise, in notation $\Delta_{\pi,n}$ we suppose that π is an arbitrary permutation.

V. Functions in small number of variables

Let us give a classification of vectorial Boolean functions in $\Delta_{\pi,n}$, where $n = 2, 3, 4$ and $\pi \in S_n$.

Proposition 5.1: It holds $|\Delta_{\pi,n}| = 2^{2^n}$ for an arbitrary permutation π on n variables.

Proof: If Boolean functions f and g are not equal, then vectorial Boolean functions F_π and G_π are not equal, because their first coordinates are f and g respectively. So $|\Delta_{\pi,n}|$ equals to number of Boolean functions in n variables which is 2^{2^n} □

All of the following propositions in this chapter are obtained by the results of computer programs.

Let A_n be the set of all derangement permutations for n elements. For example, A_4 contains permutations: $\left(\begin{smallmatrix}1&2&3&4\\2&1&4&3\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\2&3&4&1\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\2&4&1&3\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\3&1&4&2\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\3&4&1&2\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\3&4&2&1\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\4&1&2&3\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\4&3&1&2\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&2&3&4\\4&3&2&1\end{smallmatrix}\right)$.

A. Case $n = 2$

Note that for any Boolean function f in 2 variables $nl(f) \leqslant 1$, $deg(f) \leqslant 2$.

In the table I the crypto properties for function F_π in 2 variables are presented.

Table I
Crypto properties for function F_π in 2 variables

Nonlinearity	For any permutation $\pi \in S_2$ and any Boolean function f in 2 variables it holds $nl(F) = 0$.
Balancedness	*Proposition 5.2*
Algebraic degree	
Differential δ-uniformity	For any permutation $\pi \in S_2$ there exists a Boolean function f in 2 variables such that $\delta_F \geq 2$.

Proposition 5.2: There exist only two balanced vectorial Boolean functions F in 2 variables with $\delta_F = 2$ in $\Delta_{\rho,2}$, constructed with $f_i(x) = x_i, i = 1, 2$.

B. Case $n = 3$

For any Boolean function f in 3 variables $nl(f) \leqslant 2$, $deg(f) \leqslant 3$. In the table II the crypto properties for function F_π in 3 variables are presented.

Table II
CRYPTO PROPERTIES FOR FUNCTION F_π IN 3 VARIABLES

Nonlinearity	For permutations $\pi', \pi'' \in S_3$, $\pi' = \left(\begin{smallmatrix}1&2&3\\2&1&3\end{smallmatrix}\right)$, $\pi'' = \left(\begin{smallmatrix}1&2&3\\3&1&2\end{smallmatrix}\right)$ there exists a Boolean function f in 3 variables such that $nl(f) = nl(F_{\pi'}) = nl(F_{\pi''}) = 2$. It also holds that if $\pi \neq \pi', \pi''$ then for arbitrary Boolean function f in 3 variables $nl(F_\pi) < 2$.
Balancedness	For any permutation $\pi \in A_3$ there exists a balanced Boolean function f in 3 variables such that vectorial Boolean function F_π is balanced.
Algebraic degree	For any permutation $\pi \in S_3$ there exists a Boolean function f in 3 variables such that $deg(f) = deg(F_\pi) = 3$.
Differential δ-uniformity	For any permutation $\pi \in A_3$ there exists a Boolean function f in 3 variables such that $\delta_{F_\pi} = 2$. It also holds that if $\pi \notin A_3$ then for arbitrary Boolean function f in 3 variables $\delta_{F_\pi} \geqslant 4$.
	Proposition 5.3

Proposition 5.3: For permutations $\pi', \pi'' \in A_3$ there exists a Boolean function f in 3 variables such that if $\delta_{F\pi'} = 2$ then $\delta_{F_{\pi''}} = 2$.

C. Case $n = 4$

For any Boolean function f in 4 variables $nl(f) \leqslant 6$, $deg(f) \leqslant 4$. $|S_4| = 4! = 24$, $|A_4| = 9$. Denote as A_4^1 the subset of deranged permutations of all 4 elements, consisting of three pairs of permutations such that $\pi_{2_i}^{-1} = \pi_{1i}, i = 1, 2, 3$: $\pi_{1_1} = \left(\begin{smallmatrix}1&2&3&4\\2&3&4&1\end{smallmatrix}\right)$, $\pi_{2_1} = \left(\begin{smallmatrix}1&2&3&4\\4&1&2&3\end{smallmatrix}\right)$; $\pi_{1_2} = \left(\begin{smallmatrix}1&2&3&4\\2&4&1&3\end{smallmatrix}\right)$, $\pi_{2_2} = \left(\begin{smallmatrix}1&2&3&4\\3&1&4&2\end{smallmatrix}\right)$; $\pi_{1_3} = \left(\begin{smallmatrix}1&2&3&4\\3&4&2&1\end{smallmatrix}\right)$, $\pi_{2_3} = \left(\begin{smallmatrix}1&2&3&4\\4&3&1&2\end{smallmatrix}\right)$.

In the table III the crypto properties for function F_π in 4 variables are presented.

Table III
CRYPTO PROPERTIES FOR FUNCTION F_π IN 4 VARIABLES

Nonlinearity	There is no such vectorial Boolean function F in $\Delta_{\pi,4}$ that $nl(F) = 6$ for any permutation $\pi \in S_4$.
	Proposition 5.4
Balancedness	For any permutation $\pi \in A_4^1$ and a balanced Boolean function f in 4 variables such that $\delta_{F_\pi} = 2$, F_π is not balanced.
Algebraic degree	*Proposition 5.4*
Differential δ-uniformity	For any permutation $\pi \in A_4^1$ there exists a Boolean function f in 4 variables such that $\delta_F = 2$. It holds also if $\pi \notin A_4^1$ then for arbitrary Boolean function f in 4 variables $\delta_{F_\pi} \geqslant 4$.

Proposition 5.4: For any permutation $\pi \in A_4^1$ there exists a Boolean function f in 4 variables such that if $\delta_{F_\pi} = 2$ and nonlinearity and algebraic degree of f and F_π are the same then $\delta_{F_{\pi^{-1}}} = 2$. Moreover, nonlinearity and algebraic degree of $F_{\pi^{-1}}$ and f coincide.

VI. ONE-TO-ONE PROPERTY

In this section we study properties of a Boolean function f that provide the bijectiveness of the corresponding vectorial function F_π.

Proposition 6.1: Let $\pi \in S_n$, $F_\pi \in \Delta_{\pi,n}$. Then $F_\pi(\pi(x)) = \rho^{-1}(F_\pi(x))$ for all $x \in \mathbb{F}_2^n$;

Proof: We obtain the result directly from definition. □

Let π be an arbitrary permutation, we define action of π on $\mathbb{F}_2^n$ by the rule: if $x \in \mathbb{F}_2^n$ then $x \circ \pi = \pi(x)$. This action splits $\mathbb{F}_2^n$ into orbits relatively to π . If x in some orbit o, we call x *generator of* o. We call $O_\pi(x)$ *the orbit relative to the action of* π.

Example: Let us give some examples of the orbits.

For $n = 3$ the set $\mathbb{F}_2^n$ is divided into four orbits with respect to permutation ρ:

$O_\rho((0,0,0))$	$(0,0,0)$
$O_\rho((1,0,0))$	$(1,0,0), (0,1,0), (0,0,1)$
$O_\rho((0,1,1))$	$(1,0,1), (1,1,0), (0,1,1)$
$O_\rho((1,1,1))$	$(1,1,1)$

For $n = 4$:

$O_\rho((0,0,0,0))$	$(0,0,0,0)$
$O_\rho((1,0,0,0))$	$(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$
$O_\rho((1,0,1,0))$	$(1,0,1,0), (0,1,0,1)$
$O_\rho((1,0,0,1))$	$(1,0,0,1), (1,1,0,0), (0,1,1,0), (0,0,1,1)$
$O_\rho((0,1,1,1))$	$(0,1,1,1), (1,0,1,1), (1,1,0,1), (1,1,1,0)$
$O_\rho((1,1,1,1))$	$(1,1,1,1)$

We denote by $\Theta_{\pi,n}$ a set of all orbits relatively action of π on $\mathbb{F}_2^n$. The proposition 6.1 implies that for arbitrary $F_\pi \in \Delta_{\pi,n}$ values of elements of some π-orbit $g \in \Theta_{\pi,n}$ are elements of some ρ-orbit $q \in \Theta_{\rho,n}$, since $F(\pi^n(x)) = \rho^{-n}(F(x))$. Let $M_{\pi,n}^k = \{g \in \Theta_{\pi,n} : |g| = k\}$.

Denote $\Psi_{F_\pi,n} : \Theta_{\pi,n} \to \Theta_{\rho,n}$ obtained by the rule: $\Psi_{F_\pi,n}(O_\pi(x)) = O_\rho(F_\pi(x))$. Now we can formulate conditions of F_π one-to-one property in terms of $\Psi_{F_\pi,n}$.

Proposition 6.2: $F_\pi \in \Delta_{\pi,n}$ is an one-to-one function if and only if $\Psi_{F_\pi,n}$ is one-to-one function. In case if $\Psi_{F_\pi,n}$ is one-to-one then $|\Psi_{F_\pi,n}(g)| = |g|$, for all $g \in \Theta_{\pi,n}$.

Proof: $\Longrightarrow$ Suppose that F_π is one-to-one function and $\Psi_{F_\pi,n}$ is not. Let us note that from proposition 6.1 we have $F_\pi(\pi^k(x)) = \rho^{-k}(F_\pi(x))$, so $\Psi_{F_\pi,n}$ may map orbit of length k only to orbit of length q where $q \leqslant k$. Since if $q > k$ then $F_\pi(x) = F_\pi(\pi^k(x)) = \rho^k(F_\pi(x)) \neq F_\pi(x)$. Thus, there exists $g \in \Theta_{\rho,n}$ which is not a value of $\Psi_{F_\pi,n}$, or there exist $q, g \in \Theta_{\pi,n}$ such as $\Psi_{F_\pi,n}(q) = \Psi_{F_\pi,n}(g)$. Let us consider the first case. We have elements of g that are not values of F_π and this contradicts to the one-to-one property of F_π. The second case leads us to a contradiction because orbits $q, g \in \Theta_{\pi,n}$ are

mapped to $o \in \Theta_{\rho,n}$ and $|o| \leqslant |q|$, $|o| \leqslant |g|$. So, F_π can not be an one-to-one function.

$\Longleftarrow$ In order to prove reverse statement, let us show that if $\Psi_{F_\pi,n}$ is bijective, then the length of $\Psi_{F_\pi,n}(g)$ equals to the length of g. Consider some orbit g_1 of length k that is mapped to orbit g_2 of length q and $q < k$, then reverse function of $\Psi_{F_\pi,n}$ can not map g_2 to g_1, since $q < k$. So now we conclude that F_π is an one-to-one function, since the cardinality of the image equals to the cardinality of the preimage. □

As a consequence of proposition 6.2 we get the following result:

Proposition 6.3: If it holds for some k : $|M^k_{\pi,n}| \neq |M^k_{\rho,n}|$ then the set of one-to-one functions from $\Delta_{\pi,n}$ is empty.

Example: Let π be a partial derangement. This means π reorders elements except for some of them, which are called *fixed points*. Without loss of generality, we assume that the first coordinate is fixed. Then vectors $(1, 0, \ldots, 0), (0, \ldots, 0), (1, \ldots, 1)$ form orbits of length 1. But for ρ there are always only two orbits which lengths are equal to one: $(0, \ldots, 0)$ and $(1, \ldots, 1)$. Thus, for every π with fixed points there are no one-to-one functions from $\Delta_{\pi,n}$.

Proposition 6.2 means that in order to construct one-to-one functions $F_\pi \in \Delta_{\pi,n}$ we can use bijective maps $\Psi_n : \Theta_{\pi,n} \to \Theta_{\rho,n}$, which satisfy $|\Psi_n(g)| = |g|$, where $g \in \Theta_{\pi,n}$. Then, depending on them, we can construct $F_\pi \in \Delta_{\pi,n}$ such that $\Psi_{F_\pi,n} \equiv \Psi_n$.

Proposition 6.4: Let $\Psi_n : \Theta_{\pi,n} \to \Theta_{\rho,n}$ which satisfies $|\Psi_n(g)| = |g|$ for all $g \in \Theta_{\pi,n}$. Then restriction of Ψ_n on $M^k_{\pi,n}$ is arbitrary permutation from $S_{|M^k_{\pi,n}|}$.

Now consider the case $\pi = \rho$. We define $M^k_n = M^k_{\rho,n}$. Consider an one-to-one function Ψ_n which satisfies $|\Psi_n(g)| = |g|$ for all $g \in \Theta_{\pi,n}$. Let us construct function $F_\rho \in \Delta_{\rho,n}$ based on Ψ_n. Let $o \in \Theta_{\rho,n}$ be an orbit of length k. If value of F_ρ for some $x \in O$ is determined then value of F_ρ is determined for all $x \in F_\rho$, since $F_\rho(\rho^n(x)) = \rho^{-n}(F_\rho(x))$. Thus for every $\Psi_{F_\rho,n}$ we are able to construct $\prod_{k:k|n} k^{|M^k_n|}$ functions and they are all pairwise different.

Proposition 6.5: It holds $2^k = \sum_{\ell:\ell|k} \ell \cdot |M^\ell_n|$.

Proof: Orbits do not intersect by definition and their union give us $\mathbb{F}^n_2$. □

This formula allows us to calculate $|M^k_n|$ for every k. There are always only two orbits which lengths are equal to one, so we can calculate $|M^k_n|$ for every prime k. Then we can calculate it for every k. Therefore we can get a number of one-to-one functions from Δ_n via the following result:

Theorem 6.6: The number of one-to-one vectorial Boolean functions in class $\Delta_{\rho,n}$ is equal to $\prod_{k:k|n} |M^k_n|! \cdot k^{|M^k_n|}$.

Proof: Let us count the number of one-to-one maps Ψ_n from $\Theta_{\rho,n}$ to $\Theta_{\rho,n}$ that satisfies $|\Psi_n(g)| = |g|$, $g \in \Theta_{\rho,n}$. This map permutes elements of M^k_n, so the number of one-to-one Ψ_n is $\prod_{k:k|n} |M^k_n|!$. Every Ψ_n generates product of $\prod_{k:k|n} k^{|M^k_n|}$ different functions and combining these possibilities we obtain the result. □

VII. Conclusion

In the paper, we proposed a simple method of constructing S-boxes using Boolean functions in a small number of variables. We take a Boolean function f and a permutation π and construct a vectorial Boolean function F_π. We analyzed what cryptographic properties of F_π can be provided with respect to f and π. For arbitrary n we study whether there exist one-to-one functions in $\Delta_{\pi,n}$ and found their exact number. Also we describe one-to-one functions $F_\pi \in \Delta_{\pi,n}$ as mappings of orbits with respect to action π on $\mathbb{F}^n_2$. As a consequence, we offer an algorithm for constructing one-to-one functions that belong to $\Delta_{\pi,n}$. Other cryptographic properties of F and a larger number of variables should be studied in the future.

References

[1] "Cryptographic Protection for Data Processing System", GOST 28147-89, Gosudarstvennyi Standard of USSR, Government Committee of the USSR for Standards, 1989. (In Russian) https://tools.ietf.org/html/rfc5830

[2] A. Bogdanov et al. "PRESENT: An Ultra-Lightweight Block Cipher," Cryptographic Hardware and Embedded Systems - CHES 2007 Lecture Notes in Computer Science, pp. 450–466.

[3] J. Daemen, V. Rijmen The Design of Rijndael: AES — Advanced Encryption Standard. Berlin: Springer, 2002.

[4] "Information technology. Cryptographic data security. Block ciphers", GOST R 34.12-2015, Federal Agency on Technical Regulating and Metrology, 2015. https://tools.ietf.org/html/rfc7801

[5] Cusick T. W., Stănică P. Cryptographic Boolean Functions and Applications. USA: Acad. Press. Elsevier, 2009.

[6] C. Carlet, "Vectorial Boolean Functions for Cryptography," in Boolean Models and Methods in Mathematics, Computer Science, and Engineering, Y. Crama and P. L. Hammer, Eds. Cambridge: Cambridge University Press, 2010, pp. 398–470.

[7] C. Carlet, "Boolean Functions for Cryptography and Error-Correcting Codes," in Boolean Models and Methods in Mathematics, Computer Science, and Engineering, Y. Crama and P. L. Hammer, Eds. Cambridge: Cambridge University Press, 2010, pp. 257–397.

On properties of a bent function secondary construction

Nikolay Kolomeec

Sobolev Institute of Mathematics, Novosibirsk, Russia

Novosibirsk State University, Novosibirsk, Russia

Laboratory of Cryptography JetBrains Research

Abstract

In this work properties of a secondary bent function construction, that inverts values of the given bent function on an affine subspace, are obtained. Some results regarding normal and weakly normal bent functions are generalized. Bent functions and their dual functions are considered in the construction context.

1 Preliminaries

Let us recall some definitions. *A bent function* is a Boolean function in even number of variables that are at the maximal possible Hamming distance from the set of all affine Boolean functions. They were introduced by O. Rothaus [1]. Additional information regarding bent functions can be found in [2, 3]. $D_\alpha f$ is *the derivative* of f *in the direction* α. Let $\langle x, y\rangle = x_1y_1 \oplus \ldots \oplus x_ny_n$, where $x, y \in \mathbb{F}_2^n$. Denote by Ind_S characteristic function of a set $S \subseteq \mathbb{F}_2^n$. Let us define for $x \in \mathbb{F}_2^n$ and $k \leq n$

$$\begin{aligned} Proj_k(x) &= (x_1, \ldots, x_k), \\ Proj_k(S) &= \{Proj_k(x) \mid x \in S\}, \\ Elem_k(S) &= \{x \in \mathbb{F}_2^k \mid (x, \underbrace{0, \ldots, 0}_{n-k}) \in S\}. \end{aligned}$$

Hereinafter suppose that n is even. By $\mathcal{B}_n$ we denote the set all bent functions in n variables, by $\widetilde{f}$ — *a dual* bent function for $f \in \mathcal{B}_n$.

In this work we consider properties of bent function construction

$$f \oplus Ind_U,$$

where $f \in \mathcal{B}_n$ is a given bent function and U is an affine subspace of an arbitrary dimension. Necessary and sufficient conditions for $f \oplus Ind_U$ to be a bent function were proven by C. Carlet [4].

Theorem 1.1 (C. Carlet, 1994) *Let $f \in \mathcal{B}_n$, $L \leq \mathbb{F}_2^n$ be a linear subspace and $a \in \mathbb{F}_2^n$. Then $f \oplus Ind_{a\oplus L}$ is a bent function if and only if any of the following equivalent conditions hold:*

- *$D_\alpha f$ is balanced on $a \oplus L$ for all $\alpha \in \mathbb{F}_2^n \setminus L$;*
- *$\widetilde{f}(x) \oplus \langle a, x\rangle$ is either constant or balanced on each coset of $L^\perp$.*

We will use the second condition. Thus, the next two sections in some sense describe properties of dual bent function $\widetilde{f}$.

2 A balanced representation

Let us introduce the following notion for our convenience.

Definition 2.1 *A Boolean function f in n variables has a balanced representation by a linear subspace $L \leq \mathbb{F}_2^n$ if f is either constant or balanced on each coset of L.*

Note that any function has a balanced representation by the 0-dimensional linear subspace ("either constant or balanced" case allows us to ignore its odd cardinality). The same situation is for an 1-dimensional linear subspace.

First of all, there are some additional details regarding balanced representations of bent functions.

Theorem 2.2 *Let $f \in \mathcal{B}_n$ and L be a linear subspace, $\dim L \leq n/2$. Then*

- *f has a balanced representation by L if and only if f is constant on each of some $2^{n-2\dim L}$ different cosets of L;*
- *f can not be constant on more than $2^{n-2\dim L}$ different cosets of L.*

Note that the case $\dim L = n/2$ is very interesting for bent functions: for instance, H. Dobbertin introduced a large class of normal bent functions for this representation [5].

3 A balanced representation of iterative constructed functions

Let us consider the simplest iterative construction of bent function f_{+2} by $f \in \mathcal{B}_n$:

$$f_{+2}(x_1, \ldots, x_{n+2}) = f(x_1, \ldots, x_n) \oplus x_{n+1}x_{n+2}.$$

Recall that $f_{+2} \in \mathcal{B}_{n+2}$ if and only if $f \in \mathcal{B}_n$. Also, it is true

$$\widetilde{f_{+2}}(x_1, \ldots, x_{n+2}) = \widetilde{f}(x_1, \ldots, x_n) \oplus x_{n+1}x_{n+2}.$$

The question is the balanced representations for f and f_{+2} are connected or not. Let us prove the following proposition.

Proposition 3.1 *Let $f \in \mathcal{B}_n$ and have a balanced representation by $L \leq \mathbb{F}_2^n$. Then bent function f_{+2} has balanced representations by*

- *$L_0 = \{(x, 0, 0) \mid x \in L\}$, i. e. $\dim L_0 = \dim L$;*
- *$L_1 = \{(x, y, 0) \mid x \in L, y \in \mathbb{F}_2\}$, i. e. $\dim L_1 = \dim L + 1$.*

Moreover, there is "feedback" from the f_{+2} to f.

Theorem 3.2 *Let $f \in \mathcal{B}_n$ and f_{+2} have a balanced representation by $L \leq \mathbb{F}_2^{n+2}$. Then there exists $L' \leq \mathbb{F}_2^n$ with*

$$\dim L - 1 \leq \dim L' \leq \dim L,$$

such that f has a balanced representation by L'. Moreover, it holds

$$Elem_n(L) \leq L' \leq Proj_n(L).$$

In case $\dim L = n/2 + 1$ the theorem can be easily transform to "f is a normal if and only if f_{+2} is normal" proved in [6], i. e. it is a generalization of weakly normal and normal bent function properties.

4 Subspaces for iterative constructed functions

Using Theorem 1.1, the results of Section 3 can be generalized to the construction properties.

Proposition 4.1 *Let $f \in \mathcal{B}_n$ and $f \oplus Ind_U \in \mathcal{B}_n$, where U is an affine subspace of $\mathbb{F}_2^n$. Then for bent function f_{+2} the following holds:*

- $f_{+2} \oplus Ind_{U_1} \in \mathcal{B}_{n+2}$, *where* $U_1 = \{(x, y, 0) \mid x \in U, y \in \mathbb{F}_2\}$, *i. e.* $\dim U_1 = \dim U + 1$.
- $f_{+2} \oplus Ind_{U_2} \in \mathcal{B}_{n+2}$, *where* $U_2 = \{(x, y, z) \mid x \in U, y, z \in \mathbb{F}_2\}$, *i. e.* $\dim U_2 = \dim U + 2$;

Theorem 4.2 *Let $f_{+2} \in \mathcal{B}_{n+2}$ and $f_{+2} \oplus Ind_{a \oplus L} \in \mathcal{B}_{n+2}$, where $L \leq \mathbb{F}_2^{n+2}$, $a \in \mathbb{F}_2^{n+2}$. Then there exists $L' \leq \mathbb{F}_2^n$ with*

$$\dim L - 2 \leq \dim L' \leq \dim L - 1,$$

such that $f \oplus Ind_{Proj_n(a) \oplus L'} \in \mathcal{B}_n$. Moreover, it holds

$$Elem_n(L) \leq L' \leq Proj_n(L).$$

Similar to Theorem 3.2, in case $\dim L = n/2 + 1$ the theorem can be rephrase in terms of weakly normal bent function properties.

Here, trivial subspace dimensions for $f \in \mathcal{B}_n$ are n (just negation of the function) and $n - 1$ (addition of an affine function). So, it is naturally to skip these dimensions in the construction.

Computational experiments (see Section 5) show that for non-weakly normal bent function $f_{10} \in \mathcal{B}_{10}$ found in [7] (Fact 14) the following fact holds.

Fact 4.3 *For any affine subspace $U \leq \mathbb{F}_2^{10}$, $\dim U \leq 8$, it is true that $f_{10} \oplus Ind_U \notin \mathcal{B}_{10}$.*

Corollary 4.4 *For any $n \geq 10$ there exists a bent function $f \in \mathcal{B}_n$ such that $f \oplus \mathrm{Ind}_U \notin \mathcal{B}_n$ for any affine subspace $U \leq \mathbb{F}_2^n$ of dimension at most $n/2 + 3$.*

5 Search subspaces

For the given $f \in \mathcal{B}_n$, the algorithm described in [6] can help to construct all affine subspaces $U \leq \mathbb{F}_2^n$ (of an arbitrary dimension), such that $f \oplus Ind_U \in \mathcal{B}_n$. Though it constructs affine subspaces such that f is affine on each of them, it "sorts" cosets for convenient usage in a balanced representation.

Its complexity can be calculated in the following way:

$$n \sum_{m=1}^{n/2} \left(|L_m(\widetilde{f})| + (2^m - 2)|L_m^0(\widetilde{f})| \right) + \mathcal{O}(n2^n),$$

where $L_m(f)$ ($L_m^0(f)$) are all m-dimensional affine subspaces such that f is affine (constant) on them.

6 Count of the constructed functions

Let us define for $f \in \mathcal{B}_n$ and $0 \leq m \leq n$

$$Constr_m(f) = \{f \oplus Ind_U \mid U \text{ is an } m\text{-dimensional affine subspace of } \mathbb{F}_2^n\} \cap \mathcal{B}_n.$$

Theorem 6.1 *Let $f \in \mathcal{B}_n$ and $f \oplus Ind_U \in \mathcal{B}_n$, where U is an affine subspace of $\mathbb{F}_2^n$ of dimension at most $n/2 + 1$. Then*

$$\mathrm{supp}\{\widetilde{f} \oplus \widetilde{(f \oplus Ind_U)}\}$$

is an affine subspace too.

Corollary 6.2 *$|Constr_m(f)| = |Constr_m(\widetilde{f})|$ for $m \le n/2+1$.*

Unlike $n/2$ and $n/2+1$ dimensions, for other cases we have

- $\mathrm{supp}\{\widetilde{f} \oplus \widetilde{(f \oplus Ind_U)}\}$ may not be an affine subspace;
- $|Constr_m(f)|$ and $|Constr_m(\widetilde{f})|$ may not be equal; such bent functions in 8 variables exist, for instance, in Maiorana–McFarland class [8].

So, for an arbitrary subspace dimensions some construction properties differ from the case $m = n/2$.

It is well known that $|Constr_m(f)| = 0$ for $m < n/2$. The following theorem estimates cardinalities of all other $Constr_m(f)$.

Theorem 6.3 *For $f \in \mathcal{B}_n$ and $m \ge n/2$ it holds*

$$|Constr_m(f)| \le 2^{n-m} \prod_{i=1}^{n-m} \frac{2^{2m+2i-n}-1}{2^i-1}.$$

Moreover, for $m \le n-2$ the bound is reached if and only if f is quadratic.

This estimate generalizes the bound from [9] for the case $m = n/2$.

Acknowledgement

The work is supported by Mathematical Center in Akademgorodok, the agreement with Ministry of Science and High Education of the Russian Federation number 075–15–2019–1613, and by the Russian Foundation for Basic Research (project number 20–31–70043).

References

[1] O. Rothaus *On bent functions.* J. Combin. Theory. Ser. A, 20(3), 300–305, 1976.

[2] O. A. Logachev, A. A. Salnikov, V. V. Yashchenko *Boolean Functions in Coding Theory and Cryptography.* American Mathematical Society, 2012.

[3] N. Tokareva *Bent Functions, Results and Applications to Cryptography.* Acad. Press. Elsevier, 2015.

[4] C. Carlet *Two new classes of bent functions.* LNCS, 765, 77–101, 1994.

[5] H. Dobbertin *Construction of bent functions and balanced Boolean functions with high nonlinearity.* LNCS, 1008, 61–74, 1995.

[6] A. Canteaut, M. Daum, H. Dobbertin, G. Leander. *Finding nonnormal bent functions.* Discrete Appl. Math., 154(2), 202–218, 2006.

[7] G. Leander, G. McGuire *Construction of bent functions from near-bent functions.* J. Combin. Theory. Ser. A, 116(4), 960–970, 2009.

[8] R. L. McFarland *A family of difference sets in non-cyclic groups* J. Combin. Theory. Ser. A, 15, 1–10, 1973.

[9] N. Kolomeec *The graph of minimal distances of bent functions and its properties.* Designs, Codes and Cryptography, 85(3), 395–410, 2017.

Metrical properties of self-dual generalized bent functions

Kutsenko Aleksandr*

Sobolev Institute of Mathematics, Novosibirsk, Russia

Novosibirsk State University, Novosibirsk, Russia

Abstract

Bent functions of the form $\mathbb{F}_2^n \to \mathbb{Z}_q$ (K.-U Schmidt, 2006) are known as generalized bent (gbent) functions. In this paper we explore self-dual generalized bent functions and their metrical properties. Necessary and sufficient conditions for self-duality of Maiorana–McFarland gbent functions are given. We find the complete Hamming and Lee distance spectrums between self-dual Maiorana–McFarland gbent functions. It is proved that the set of quaternary self-dual gbent functions is metrically regular within the Lee distance. Minimal Lee distance between mentioned functions is obtained as a corollary. We define the action of a linear operator $\mathbb{C}^{2^n} \to \mathbb{C}^{2^n}$ on the set of generalized Boolean functions in n variables and characterize all unitary operators which transform the set of all generalized Boolean functions in n variables into itself and preserve self-duality. It is proved that there is no such unitary operator which assigns to every regular generalized bent function in even number n of variables its dual gbent function. In particular, from this result it follows that there is no isometric mapping of the set of all Boolean functions into itself which assigns to every bent function its dual function.

Let $\mathbb{F}_2^n$ be a set of binary vectors of length n. For $x, y \in \mathbb{F}_2^n$ denote $\langle x, y\rangle = \bigoplus_{i=1}^{n} x_i y_i$, where the sign $\oplus$ denotes a sum modulo 2.

A *generalized Boolean function* f in n variables is any map from $\mathbb{F}_2^n$ to $\mathbb{Z}_q$, the integers modulo q. The set of generalized Boolean functions in n variables is denoted by $\mathcal{GF}_n^q$. Let $\omega = e^{2\pi i/q}$. A *sign* function of $f \in \mathcal{GF}_n^q$ is a complex valued function ω^f, we will also refer to it as to a complex vector $\left(\omega^{f_0}, \omega^{f_1}, ..., \omega^{f_{2^n-1}}\right)$ of length 2^n, where $(f_0, f_1, ..., f_{2^n-1})$ is a vector of values of the function f.

The *Hamming weight* $\mathrm{wt}_H(x)$ of the vector $x \in \mathbb{F}_2^n$ is the number of nonzero coordinates of x. The *Hamming distance* $\mathrm{dist}_H(f, g)$ between generalized Boolean functions f, g in n variables is the cardinality of the set $\{x \in \mathbb{F}_2^n | f(x) \neq g(x)\}$. The Lee weight of the element $x \in \mathbb{Z}_q$ is $\mathrm{wt}_L(x) = \min\{x, q - x\}$. The Lee distance $\mathrm{d}_L(f, g)$ between $f, g \in \mathcal{GF}_n^q$ is

$$\mathrm{dist}_L(f, g) = \sum_{x \in \mathbb{F}_2^n} \mathrm{wt}_L\left(\delta(x)\right),$$

where $\delta \in \mathcal{GF}_n^q$ and $\delta(x) = f(x) + (q-1)g(x)$ for any $x \in \mathbb{F}_2^n$. For Boolean case $q = 2$ the Hamming distance coincides with the Lee distance.

The *(generalized) Walsh–Hadamard transform* of $f \in \mathcal{GF}_n^q$ is the complex valued function:

$$\mathcal{H}_f(y) = \sum_{x \in \mathbb{F}_2^n} \omega^{f(x)} (-1)^{\langle x, y\rangle}.$$

A generalized Boolean function f in n variables is said to be *generalized bent* (gbent) if

$$|\mathcal{H}_f(y)| = 2^{n/2},$$

*The work was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by Russian Foundation for Basic Research (project no. 18-07-01394, 20-31-70043) and Laboratory of Cryptography JetBrains Research.

for all $y \in \mathbb{F}_2^n$ [9]. If there exists such $\widetilde{f} \in \mathcal{GF}_n^q$ that $\mathcal{H}_f(y) = \omega^{\widetilde{f}(y)} 2^{n/2}$ for any $y \in \mathbb{F}_2^n$, the gbent function f is said to be *regular* and $\widetilde{f}$ is called its *dual*. Note that $\widetilde{f}$ is generalized bent as well. A regular gbent function f in said to be *self-dual* if $f = \widetilde{f}$, and *anti-self-dual* if $f = \widetilde{f} + \frac{q}{2}$. Consequently, it is the case only for even q. So throughout this paper we assume that q is a natural even number.

In paper [6] for the case when q is divisible by 4, necessary and sufficient conditions for the bentness of generalized Boolean functions of the form

$$f(x) = \sum_{i=1}^{n} \lambda_i x_i + \lambda_0,$$

where $\lambda_0, \lambda_1, ..., \lambda_n \in \mathbb{Z}_q$, were obtained. Functions from this class are referred to as *affine* functions.

It is well known that Boolean bent function and, as a consequence, self-dual Boolean bent function can not be affine. The next result shows non-existence of self-dual generalized bent functions in the class of affine functions.

Theorem 0.1 *There are no self-dual generalized bent functions in n variables of the form*

$$f(x) = \sum_{i=1}^{n} \lambda_i x_i + \lambda_0,$$

where $\lambda_0, \lambda_1, ..., \lambda_n \in \mathbb{Z}_q$.

Bent functions in $2k$ variables which have a representation

$$f(x, y) = \langle x, \pi(y) \rangle \oplus g(y),$$

where $x, y \in \mathbb{F}_2^k$, $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and g is a Boolean function in k variables, form the well known *Maiorana–McFarland* class of bent functions. It is known [1] that a dual of a Maiorana–McFarland bent function $f(x, y)$ is equal to

$$\widetilde{f}(x, y) = \langle \pi^{-1}(x), y \rangle \oplus g\left(\pi^{-1}(x)\right).$$

A generalization of this construction for the case $q = 4$ was given by Schmidt in [9]. In [11] this construction was given for any even q, thus, forming the following construction

$$f(x, y) = \frac{q}{2} \langle x, \pi(y) \rangle \oplus g(y),$$

where $x, y \in \mathbb{F}_2^k$, $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and g is a generalized Boolean function in k variables. Its dual is

$$\widetilde{f}(x, y) = \frac{q}{2} \langle \pi^{-1}(x), y \rangle \oplus g\left(\pi^{-1}(x)\right).$$

In the article [2] necessary and sufficient conditions of (anti-)self-duality of Maiorana–McFarland bent functions, denoted by $\mathrm{SB}^+_{\mathcal{M}}(n)$ ($\mathrm{SB}^-_{\mathcal{M}}(n)$), were given. In [10] quaternary self-dual Maiorana–McFarland bent functions were studied and necessary and sufficient conditions of self-duality were obtained.

In the current work we generalize these results for any even q. Denote the sets of (anti-)self-dual generalized Maiorana–McFarland bent functions by $\mathrm{SB}^+_{\mathcal{M}^q}(n)$ ($\mathrm{SB}^-_{\mathcal{M}^q}(n)$)

Theorem 0.2 *A generalized Maiorana–McFarland bent function*

$$f(x, y) = \frac{q}{2} \langle x, \pi(y) \rangle + g(y), \ x, y \in \mathbb{F}_2^{n/2},$$

is (anti-)self-dual bent if and only if for any $y \in \mathbb{F}_2^{n/2}$

$$\pi(y) = L\left(y \oplus b\right), \quad g(y) = \frac{q}{2} \langle b, y \rangle + d,$$

where $L \in \mathcal{O}_{n/2}, b \in \mathbb{F}_2^{n/2}$, $\mathrm{wt}\,(b)$ is even (odd), $d \in \mathbb{Z}_q$.

It follows that the number of such functions is a function of q and the cardinality of the orthogonal group.

Corollary 0.3 *It holds*

$$\left|\mathrm{SB}^{+}_{\mathcal{GM}^q}(n)\right| = \left|\mathrm{SB}^{-}_{\mathcal{GM}^q}(n)\right| = q \cdot 2^{n/2-1} \left|\mathcal{O}\left(n/2, \mathbb{F}_2\right)\right|.$$

In paper [3] the possible Hamming distances between (anti-)self-dual Maiorana–McFarland bent functions for the Boolean case were studied and the complete Hamming distances of these distances was presented, namely it was shown that for $f, g \in \mathrm{SB}^{+}_{\mathcal{M}}(n) \cup \mathrm{SB}^{-}_{\mathcal{M}}(n)$, then

$$\mathrm{dist}(f,g) \in \left\{2^{n-1}, 2^{n-1}\left(1 \pm \frac{1}{2^r}\right), r = 0, 1, ..., n/2-1\right\}.$$

Moreover, it was shown that if either $f, g \in \mathrm{SB}^{+}_{\mathcal{M}}(n)$ or $f, g \in \mathrm{SB}^{-}_{\mathcal{M}}(n)$, then all distances given above are attainable. If f is self-dual bent and g is anti-self-dual bent, then $\mathrm{dist}(f,g) = 2^{n-1}$.

In the current work we generalize this result for any even q in both Hamming and Lee distances. Denote the mentioned spectrum for the Hamming distance by $\mathrm{Sp}_H\left(\mathrm{SB}^{+}_{\mathcal{GM}^q}(n) \cup \mathrm{SB}^{-}_{\mathcal{GM}^q}(n)\right)$, while for the Lee distance the notation $\mathrm{Sp}_L\left(\mathrm{SB}^{+}_{\mathcal{GM}^q}(n) \cup \mathrm{SB}^{-}_{\mathcal{GM}^q}(n)\right)$ is used.

Theorem 0.4 *It holds*

$$\mathrm{Sp}_H\left(\mathrm{SB}^{+}_{\mathcal{GM}^q}(n) \cup \mathrm{SB}^{-}_{\mathcal{GM}^q}(n)\right) = \left\{2^{n-1}\right\} \cup \bigcup_{r=0}^{n/2-1} \left\{2^{n-1}\left(1 \pm \frac{1}{2^r}\right)\right\},$$

$$\mathrm{Sp}_L\left(\mathrm{SB}^{+}_{\mathcal{GM}^q}(n) \cup \mathrm{SB}^{-}_{\mathcal{GM}^q}(n)\right) = \left\{q \cdot 2^{n-2}\right\} \cup \bigcup_{\mathrm{w}=0}^{q/2} \bigcup_{r=0}^{n/2-1} \left\{q \cdot 2^{n-2}\left(1 \pm \frac{1}{2^r}\right) \mp \mathrm{w} \cdot 2^{n-r}\right\}.$$

Moreover, all given distances are attainable.

It is possible to derive minimal distances from these spectrums.

Corollary 0.5 *The minimal Lee distance between generalized (anti-)self-dual Maiorana–McFarland bent functions in n variables is equal to $2^{n-3}q$, while the minimal Hamming distance is 2^{n-2}.*

Recall that $\mathrm{RM}(r,m)$ is the length 2^m linear code over that is generated by the monomials of order at most r in variables $x_1, x_2, ..., x_m$, its minimal Lee distance is equal to 2^{m-r} [8]. Hence for $\mathrm{RM}(2,m)$ minimal Lee distance is equal to 2^{n-2}. From the obtained results it follows that

Corollary 0.6 *The minimal Lee distance 2^{n-2} between quadratic (generalized) bent functions is attainable on (anti-)self-dual Maiorana–McFarland bent functions from $\mathcal{GM}^q_n$ only for $q = 2$.*

Let $X \subseteq \mathbb{Z}_q^n$ be an arbitrary set and let $y \in \mathbb{Z}_q^n$ be an arbitrary vector. Define the *distance* between y and X as $\mathrm{dist}(y, X) = \min\limits_{x \in X} \mathrm{dist}(y, x)$. The *maximal distance* from the set X is

$$\mathrm{d}(X) = \max_{y \in \mathbb{Z}_q^n} \mathrm{dist}(y, X).$$

In coding theory this number is also known as the *covering radius* of the set X. A vector $z \in \mathbb{Z}_q^n$ is called *maximally distant* from a set X if $\mathrm{dist}(z, X) = \mathrm{d}(X)$. The set of all maximally distant vectors from the set X is called the *metrical complement* of the set X and denoted by $\widehat{X}$. A set X is said to be *metrically regular* if $\widehat{\widehat{X}} = X$. A subset of Boolean functions is said to be *metrically regular* if the set of corresponding vectors of values is metrically regular [13].

In paper [4] it was proved that the set of Boolean self-dual bent functions is metrically regular within the Hamming distance. In current work we prove that within Lee distance this statement holds for the quaternary case $q = 4$ as well.

Theorem 0.7 *The sets of (anti-)self-dual generalized quaternary bent functions are metrically regular within the Lee distance.*

Let $\varphi : \mathbb{C}^{2^n} \to \mathbb{C}^{2^n}$ be a linear operator with matrix A in canonical basis of the space $\mathbb{C}^n$. We will say that φ transforms the generalized function $f \in \mathcal{GF}_n^q$ with sign function $F = \omega^f$ to the function $f' \in \mathcal{GF}_n^q$ if the sign function of f' is equal to AF. Denote by $\mathcal{U}_n^q$ the set of unitary operators $\mathbb{C}^{2^n} \to \mathbb{C}^{2^n}$ which transform the set of q-ary generalized Boolean functions in n variables into itself.

The general form of these mappings is given by the following

Theorem 0.8 *Every operator from $\mathcal{U}_n^q$ can be uniquely represented in the form*

$$f(x) \longrightarrow f\left(\pi(x)\right) + g(x),$$

where π is a permuation on $\mathbb{F}_2^n$ and $g \in \mathcal{GF}_n^q$.

Corollary 0.9 *Every operator from $\mathcal{U}_n^q$ preserves Lee and Hamming distances between generalized Boolean functions and Euclidian distance between their sign functions.*

Corollary 0.10 *It holds*

$$|\mathcal{U}_n^q| = 2^n! \cdot q^{2^n}.$$

From Markov's theorem (1956) [7] it follows that the general form of isometric mappings of all Boolean functions in n variables to itself is

$$f(x) \longrightarrow f(\pi(x)) \oplus g(x),$$

where π is a permutation on the set $\mathbb{F}_2^n$ and $g \in \mathcal{F}_n$ [7]. Thus for the Boolean case we have the following:

Corollary 0.11 *For $q = 2$ there is an one-to-one correspondence between the set $\mathcal{U}_n^q$ and the set of isometric mappings of all Boolean functions in n variables into itself, defined by Markov's theorem.*

Corollary 0.12 *It holds*

$$|\mathcal{U}_n^q| = 2^n! \cdot q^{2^n}.$$

In paper [5] isometric mappings of all Boolean functions in n variables into itself which preserve self-duality were completely described, namely it was proved that isometric mapping $\varphi : f(x) \to f(\pi(x)) \oplus g(x)$ preserves self-duality if and only if

$$\pi(x) = L\left(x \oplus c\right), \ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{F}_2$. It was also shown that isometric mappings of all Boolean functions in n variables into itself preserves self-duality if and only if it preserves anti-self-duality.

We generalize this result within the set $\mathcal{U}_n^q$:

Theorem 0.13 *Isometric mapping $\varphi : f(x) \to f(\pi(x)) \oplus g(x)$ preserves (anti-)self-duality of generalized regular bent function if and only if*

$$\pi(x) = L\left(x \oplus c\right), \ x \in \mathbb{F}_2^n,$$

and

$$g(x) = \frac{q}{2} \langle c, x \rangle \oplus d, \ x \in \mathbb{F}_2^n,$$

where $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ is even, $d \in \mathbb{Z}_q$.

Regarding the mapping $f \to \widetilde{f}$, which assigns the dual bent function to every regular bent function the following statement holds.

Theorem 0.14 *If n is an even number, then in the set $\mathcal{U}_n^q$ there is no such operator which assigns the dual gbent function to every regular gbent function from the set $\mathcal{GB}_n^q$.*

Corollary 0.15 *There is no isometric mapping of the set of all Boolean functions into itself which assigns to every Boolean bent function its dual Boolean bent function.*

From this fact and the general form of isometric mappings which preserve bentness [12] it follows that

Corollary 0.16 *The mapping defined on the set of Boolean bent functions in n variables as follows*

$$f(x) \longrightarrow \widetilde{f}(x),$$

cannot be represented as a combination of an affine transform of coordinates and an affine shift.

References

[1] Carlet C. Boolean functions for cryptography and error correcting code. In: Crama Y., Hammer P.L. (eds.) Boolean Models and Methods in Mathematics, Computer Science, and Engineering. p. 257–397. Cambridge University Press, Cambridge (2010).

[2] Carlet C., Danielson L.E., Parker M.G., Solé. P., *Self-dual bent functions.* Int. J. Inform. Coding Theory, **1**, 384–399 (2010).

[3] Kutsenko A.V., *The Hamming Distance Spectrum Between Self-Dual Maiorana–McFarland Bent Functions*, Journal of Applied and Industrial Mathematics, **12**(1), 112–125 (2018).

[4] Kutsenko A., *Metrical properties of self-dual bent functions*, Des. Codes Cryptogr. **88**, 201–222 (2020).

[5] Kutsenko A., *The group of automorphisms of the set of self-dual bent functions*, Cryptology ePrint Archive: Report 2019/1408.

[6] Singh B.K., *On cross-correlation spectrum of generalized bent functions in generalized Maiorana–McFarland class.* Information Sciences Letters. **2**(3), 139–145 (2013).

[7] Markov A. A., *On transformations without error propagation.* In: Selected Works, Vol. II: Theory of Algorithms and Constructive Mathematics. Mathematical Logic. Informatics and Related Topics, p. 70–93, MTsNMO, Moscow (2003) [Russian].

[8] Paterson K.G., Jones A.E., *Efficient decoding algorithms for generalized Reed–Muller codes.* IEEE Trans. Commun., vol. 48, no. 8, pp. 1272–1285, 2000.

[9] Schmidt K.-U., *Quaternary constant-amplitude codes for multicode CDMA.* IEEE Trans. Inform. Theory, **55**, 1824–1832 (2009).

[10] Sok L., Shi M., Solé P., *Classification and Construction of quaternary self-dual bent functions.* Cryptogr. Commun. **10**, 277–289 (2018).

[11] Stănică P., Martinsen T., Gangopadhyay S., Singh B. K., *Bent and generalized bent Boolean functions.* Des. Codes Cryptogr. **69**, 77–94 (2013).

[12] Tokareva N.N., *The group of automorphisms of the set of bent functions.* Discrete Mathematics and Applications, **20**(5), 655–664 (2010).

[13] Tokareva N., Bent Functions, *Results and Applications to Cryptography.* Acad. Press. Elsevier, 2015.

Metric regularity of Reed-Muller codes *

Alexey Oblaukhov[1,2]

[1]Sobolev Institute of Mathematics, Novosibirsk, Russia
[2]Novosibirsk State University, Novosibirsk, Russia

Abstract

In this work we study metric properties of the well-known family of binary Reed-Muller codes. Let A be an arbitrary subset of the Boolean cube, and $\widehat{A}$ be the metric complement of A — the set of all vectors of the Boolean cube at the maximal possible distance from A. If the metric complement of $\widehat{A}$ coincides with A, then the set A is called a *metrically regular set*. The problem of investigating metrically regular sets appeared when studying *bent functions*, which have important applications in cryptography and coding theory and are also one of the earliest examples of a metrically regular set. In this work we describe metric complements and establish the metric regularity of the codes $\mathcal{RM}(0,m)$ and $\mathcal{RM}(k,m)$ for $k \geqslant m-3$. Additionally, the metric regularity of the codes $\mathcal{RM}(1,5)$ and $\mathcal{RM}(2,6)$ is proved. Combined with previous results by Tokareva N. (2012) concerning duality of affine and bent functions, this proves the metric regularity of most Reed-Muller codes with known covering radius. It is conjectured that all Reed-Muller codes are metrically regular.

1 Introduction

The problem of investigating and classifying *metrically regular sets* was posed by Tokareva [14, 15] when studying metric properties of *bent functions* [11]. A Boolean function f in even number of variables m is called a *bent function* if it is at the maximal possible distance from the set of affine functions.

Bent functions have various applications in cryptography, coding theory and combinatorics [6, 15]. In cryptography, bent functions are valued because of their outstanding nonlinearity, which allows one to construct S-boxes for block ciphers which possess high resistance to the linear cryptanalysis [6]. However, many problems related to bent functions remain unsolved; in particular, the gap between the best known lower and upper bound on the number of bent functions is extremely large; currently known constructions of bent functions are rather scarse. In 2012 [14], Tokareva has proved that, like bent functions are maximally distant from affine functions, affine functions are at the maximal possible distance from bent functions, thus establishing the *metric regularity* of both sets. This discovery arouses interest in studying the property of metric regularity in order to better understand the structure of the set of bent functions.

Let us briefly overview the results obtained in this area. Metric regularity of several classes of *partition set functions* is studied in [13]. The work [4] examines metric properties of self-dual bent functions. Metric regularity has been actively investigated by the author: metric complements of linear subspaces of the Boolean cube are studied in the paper [8], while the works [9] and [10] are studying possible sizes of the largest and smallest metrically regular set.

In this work we investigate metric properties of Reed-Muller codes. Among the codes of high order, covering radii of the codes $\mathcal{RM}(k,m)$, for $k \geqslant m-3$ are known. The covering radius of $\mathcal{RM}(1,m)$ for odd $m > 7$ is unknown, but has been determined for $\mathcal{RM}(1,5)$ [1] and $\mathcal{RM}(1,7)$ [7, 3]. In [12], Schatz has found the covering radius of $\mathcal{RM}(2,6)$, while recently Wang has

*The work was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by Russian Foundation for Basic Research (projects no. 18-07-01394, 19-31-90093) and Laboratory of Cryptography JetBrains Research.

established the covering radius of $\mathcal{RM}(2,7)$ [16]. For $m > 7$, the covering radius of $\mathcal{RM}(2,m)$ is still unknown. We prove that the codes $\mathcal{RM}(k,m)$, for $k = 0$ and $k \geqslant m-3$ and the codes $\mathcal{RM}(1,5)$ and $\mathcal{RM}(2,6)$ are metrically regular and also describe their metric complements in most cases.

2 Preliminaries

Let $\mathbb{F}_2^n$ be the space of binary vectors of length n with the Hamming metric. The *Hamming distance* $d(\cdot,\cdot)$ between two binary vectors is defined as the number of coordinates in which these vectors differ, while $wt(\cdot)$ denotes the *weight* of a vector, i.e. the number of nonzero values it contains. The plus sign + denotes addition modulo two (componentwise in case of vectors).

Let $X \subseteq \mathbb{F}_2^n$ be an arbitrary set and $y \in \mathbb{F}_2^n$ be an arbitrary vector. The distance from the vector y to the set X is defined as

$$d(y,X) = \min_{x \in X} d(y,x).$$

The *covering radius* of the set X is defined as

$$\rho(X) = \max_{z \in \mathbb{F}_2^n} d(z,X).$$

The set X with $\rho(X) = r$ is also called a *covering code* [2] of radius r.

Consider the set

$$Y = \{y \in \mathbb{F}_2^n | d(y,X) = \rho(X)\}$$

of all vectors at the maximal possible distance from the set X. This set is called the *metric complement* [8] of X and is denoted by $\widehat{X}$. Vectors from the metric complement are sometimes called *deep holes* of a code. If $\widehat{\widehat{X}} = X$ then the set X is said to be *metrically regular* [15].

Note that metrically regular sets always come in pairs, i.e. if A is a metrically regular set, then its metric complement $\widehat{A}$ is also a metrically regular set and both of them have the same covering radius. For some simple examples of metric complements and metrically regular sets, refer to [8, 9, 10].

The following trivial auxiliary lemma, established in [8], will be used throughout the paper.

Lemma 2.1 *Let $C \subseteq \mathbb{F}_2^n$ be a linear code. Then $\rho(\widehat{C}) = \rho(C)$ and a vector $x \in \mathbb{F}_2^n$ is in $\widehat{\widehat{C}}$ if and only if $x + \widehat{C} = \widehat{C}$.*

Let $\mathcal{F}^m$ be the set of all Boolean functions in m variables. The Reed-Muller code of order k is defined as:

$$\mathcal{RM}(k,m) = \{f \in \mathcal{F}^m : \deg(f) \leqslant k\},$$

where $\deg(\cdot)$ denotes the degree of the *algebraic normal form (ANF)* of the function.

Let f and g be two functions in m variables. Denote as $L_{\mathrm{A}}^{\mathrm{b}} : \mathbb{F}_2^m \to \mathbb{F}_2^m$ the affine transformation of the variables with the matrix A and the vector b):

$$(f \circ L_{\mathrm{A}}^{\mathrm{b}})(\mathrm{x}) = f(\mathrm{Ax} + \mathrm{b}).$$

Here $\circ$ denotes the composition of the functions. If the vector b is zero, it will be omitted from the notation. Functions f and g are called *linearly equivalent* if one can be obtained from the other by applying a nonsingular linear transformation to the variables, i.e. $f = g \circ L_{\mathrm{A}}$, where $\det \mathrm{A} \neq 0$.

Extended affine equivalence is more common when classifying boolean functions: functions f and g are called EA-*equivalent* if there exists a nonsingular linear transformation of variables A, a boolean vector b of length m and a function h of degree at most 1 such that $f = g \circ L_{\mathrm{A}}^{\mathrm{b}} + h$.

For our study we will use a variant of these two equivalence relations, which will be referred to as *extended linear equivalence (to the power of k)*. Functions f and g are called EL^k*-equivalent* if there exists a nonsingular binary matrix A and a function h of degree at most k such that

$$f = g \circ L_{\mathrm{A}} + h.$$

It is easy to see that this relation is indeed an equivalence. We will denote this equivalence by $f \overset{k}{\sim} g$.

The Reed-Muller code of order k in m variables is usually denoted as $\mathcal{RM}(k, m)$. Since we will refer to these codes regularly, we will instead often use $\mathcal{R}_{k,m}$ to denote the Reed-Muller code of order k in m variables. We will sometimes omit the number of variables m if it is clear from the context.

3 The Reed-Muller code $\mathcal{RM}(1, 5)$

In the work [1], Berlekamp and Welch presented a partition of all cosets of the $\mathcal{R}_{1,5}$ code into 48 classes with respect to the EA-equivalence and obtained weight distributions for each class of cosets. Four of these cosets contain only codewords of weight 12 and higher, and those cosets constitute the metric complement of $\mathcal{R}_{1,5}$. Thus we can present the metric complement of this code as:

$$\widehat{\mathcal{R}}_{1,5} = \{f : f \overset{\mathrm{EA}}{\sim} g \text{ for some } g \text{ from one of 4 farthest classes}\}$$

Since $\mathcal{R}_{1,5}$ is linear, it follows that $\rho(\widehat{\mathcal{R}}_{1,5}) = \rho(\mathcal{R}_{1,5}) = 12$, and $f \in \widehat{\widehat{\mathcal{R}}}_{1,5}$ if and only if $f + \widehat{\mathcal{R}}_{1,5} = \widehat{\mathcal{R}}_{1,5}$. Thus, in order to establish the metric regularity of $\mathcal{R}_{1,5}$, we must prove that for every $f \notin \mathcal{R}_{1,5}$ it holds $f + \widehat{\mathcal{R}}_{1,5} \neq \widehat{\mathcal{R}}_{1,5}$.

This is done by taking a representative f_c from every class of cosets C (aside from $\mathcal{R}_{1,5}$ itself) and showing that there exists a function $g_c \in \widehat{\mathcal{R}}_{1,5}$ such that $f_c + g_c \notin \widehat{\mathcal{R}}_{1,5}$. Since the metric complement $\widehat{\mathcal{R}}_{1,5}$ consists of EA-equivalence classes, this proves that none of the functions from the class C belong to $\widehat{\mathcal{R}}_{1,5}$. Therefore, the following holds:

Theorem 3.1 *The code $\mathcal{R}_{1,5}$ is metrically regular.*

4 The Reed-Muller codes of orders 0, m, $m-1$ and $m-2$

The Reed-Muller codes of orders 0, m and $m-1$ coincide with the repetition code, the whole space and the even weight code respectively. It is trivial that all of them are metrically regular.

The Reed-Muller code of order $m-2$ has covering radius 2 [2]. By definition, it consists of all Boolean functions of degree at most $m-2$. Since all functions of degree m have odd weight, and all functions of smaller degree have even weight, functions of degree m are at distance 1 from $\mathcal{R}_{m-2}$, while functions of degree $m-1$ are at distance 2 and therefore

$$\widehat{\mathcal{R}}_{m-2} = \mathcal{R}_{m-1} \setminus \mathcal{R}_{m-2}.$$

Since $\mathcal{R}_{m-2}$ is linear, $\rho(\widehat{\mathcal{R}}_{m-2}) = \rho(\mathcal{R}_{m-2}) = 2$ and thus functions of degree m are at distance 1 from $\widehat{\mathcal{R}}_{m-2}$. It follows that $\widehat{\widehat{\mathcal{R}}}_{m-2} = \mathcal{R}_{m-2}$ and $\mathcal{R}_{m-2}$ is metrically regular.

5 The Reed-Muller code of order $m-3$

5.1 Covering radius

McLoughlin [5] has proved that

$$\rho(\mathcal{R}_{m-3}) = \begin{cases} m+1, & \text{if } m \text{ is odd,} \\ m+2, & \text{if } m \text{ is even.} \end{cases}$$

This result is reestablished by Cohen et al in the book "Covering codes" [2], using a method of syndrome matrices, different from that in [5]. This method allows us not only to obtain covering radius of the Reed-Muller code of order $m-3$, but also to describe the metric complement of this code. As with the covering radius, the cases of even and odd m are distinct.

5.2 Case m is even

In this case, the metric complement can be described as follows:

$$\widehat{\mathcal{R}}_{m-3} = \bigcup_{g \in G} (g + \mathcal{R}_{m-3}),$$

where

$$G = \{g : \mathrm{supp}(g) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_m, \mathrm{x}_1 + \ldots + \mathrm{x}_m\}, \\ \{\mathrm{x}_1, \ldots, \mathrm{x}_m\} \text{ are linearly independent}\}.$$

It is easy to see that all functions in G form an equivalence class with respect to the linear equivalence. Let us pick any function g^* from this class. We can now say that a function g is in $\widehat{\mathcal{R}}_{m-3}$ if and only if $g = g^* \circ L_{\mathrm{A}} + h$ for some nonsingular matrix A and some function h of degree at most $m-3$, or, in other words, g is in $\widehat{\mathcal{R}}_{m-3}$ if and only if g is EL^{m-3}-equivalent to g^*. Therefore,

$$\widehat{\mathcal{R}}_{m-3} = \{g : g \overset{m-3}{\sim} g^*\},$$

where g^* is some function from the class G (or from $\widehat{\mathcal{R}}_{m-3}$, since all functions in metric complement are EL^{m-3}-equivalent).

5.3 Case m is odd

In this case, the metric complement can be described as follows:

$$\widehat{\mathcal{R}}_{m-3} = \bigcup_{g \in G_1 \cup G_2} (g + \mathcal{R}_{m-3}),$$

where

$$G_1 = \{g : \mathrm{supp}(g) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_m\}, \{\mathrm{x}_1, \ldots, \mathrm{x}_m\} \text{ are linearly independent}\},$$

and

$$G_2 = \{g : \mathrm{supp}(f) = \{0, \mathrm{x}_1, \mathrm{x}_2 \ldots, \mathrm{x}_{m-1}, \mathrm{x}_1 + \ldots + \mathrm{x}_{m-1}\}, \\ \{\mathrm{x}_1, \ldots, \mathrm{x}_{m-1}\} \text{ are linearly independent}\}.$$

Same as with the case of even m, all functions in G_1 form an equivalence class with respect to the linear equivalence, so do functions from G_2. If we now choose a representative from each class, g_1^* from G_1 and g_2^* from G_2, we can describe metric complement in the following manner:

$$\widehat{\mathcal{R}}_{m-3} = \{g : g \overset{m-3}{\sim} g_1^*\} \cup \{g : g \overset{m-3}{\sim} g_2^*\}.$$

5.4 Metric regularity

Since the code $\mathcal{R}_{m-3}$ is linear, it follows that $\rho(\widehat{\mathcal{R}}_{m-3}) = \rho(\mathcal{R}_{m-3})$ and a function f is in $\widehat{\widehat{\mathcal{R}}}_{m-3}$ if and only if $f + \widehat{\mathcal{R}}_{m-3} = \widehat{\mathcal{R}}_{m-3}$. Thus, like in the Section 3, we prove the metric regularity of $\mathcal{R}_{m-3}$ by proving that no functions other that those contained in $\mathcal{R}_{m-3}$ preserve the metric complement under addition, using the representations of metric complements obtained in the previous subsections.

6 The Reed-Muller code $\mathcal{RM}(2,6)$

Let us consider one other special case. If we change the order of values in the value vectors of functions so that the first half of values corresponds to the values of the function when the last variable is set to 0, and the other half corresponds to the values of the function when the last variable is set to 1, then each Reed-Muller code (for $m > 1$, $r > 0$) can be inductively defined as follows:

$$\mathcal{R}_{r,m} = \{(\mathbf{u}, \mathbf{u}+\mathbf{v}) | \mathbf{u} \in \mathcal{R}_{r,m-1}, \mathbf{v} \in \mathcal{R}_{r-1,m-1}\}.$$

In particular,

$$\mathcal{R}_{2,6} = \{(\mathbf{u}, \mathbf{u}+\mathbf{v}) | \mathbf{u} \in \mathcal{R}_{2,5}, \mathbf{v} \in \mathcal{R}_{1,5}\}.$$

Since both $\mathcal{R}_{2,5}$ and $\mathcal{R}_{1,5}$ were shown to be metrically regular, this construction proves useful and allows us to establish the metric regularity of the code $\mathcal{R}_{2,6}$ as well. From now on, vectors in bold will represent value vectors of functions in 5 variables (of length 32), while value vectors of 6-variable functions will be presented as pairs of value vectors of 5-variable functions.

Let $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}}+\tilde{\mathbf{v}}) \in \widehat{\widehat{\mathcal{R}}}_{2,6}$. We will prove that $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}}+\tilde{\mathbf{v}})$ is in $\mathcal{R}_{2,6}$ in two steps: first we establish that $\tilde{\mathbf{u}}$ is in $\mathcal{R}_{2,5}$, then we prove that $\tilde{\mathbf{v}}$ is in $\mathcal{R}_{1,5}$. The following results heavily rely on the fact that $\mathcal{R}_{2,6}$ attains the upper bound on the covering radius provided by the $(\mathbf{u}, \mathbf{u}+\mathbf{v})$ construction, i.e. $\rho(\mathcal{R}_{2,6}) = \rho(\mathcal{R}_{2,5}) + \rho(\mathcal{R}_{1,5})$ [12].

Recall (Section 5) that $\widehat{\mathcal{R}}_{2,5} = \{g : g \overset{2}{\sim} g_1\} \cup \{g : g \overset{2}{\sim} g_2\}$, where g_1 and g_2 are some representatives of two EL^2-equivalence classes. Let us denote

$$\widehat{\mathcal{R}}^1_{2,5} := \{g : g \overset{2}{\sim} g_1\},\ \widehat{\mathcal{R}}^2_{2,5} := \{g : g \overset{2}{\sim} g_2\}.$$

The following lemma is useful when proving that $\tilde{\mathbf{u}} \in \mathcal{R}_{2,5}$:

Lemma 6.1 *For each $i = 1, 2$ one of the following statements holds:*

1. $\forall \mathbf{y} \in \widehat{\mathcal{R}}^i_{2,5}\ \forall \mathbf{w} \in \mathbb{F}_2^{32}$ *it holds* $(\mathbf{y}, \mathbf{w}) \notin \widehat{\mathcal{R}}_{2,6}$;
2. $\forall \mathbf{y} \in \widehat{\mathcal{R}}^i_{2,5}\ \exists \mathbf{w} \in \mathbb{F}_2^{32}$ *such that* $(\mathbf{y}, \mathbf{w}) \in \widehat{\mathcal{R}}_{2,6}$;

This lemma tells us that for each EL^2-equivalence class of $\widehat{\mathcal{R}}_{2,5}$, either all vectors appear in the metric complement of $\mathcal{R}_{2,6}$ as the first half of the vector, or no vectors do. Since for any $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}}+\tilde{\mathbf{v}}) \in \widehat{\widehat{\mathcal{R}}}_{2,6}$ it holds $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}}+\tilde{\mathbf{v}}) + \widehat{\mathcal{R}}_{2,6} = \widehat{\mathcal{R}}_{2,6}$, it is easy to show that $\tilde{\mathbf{u}}$ must keep $\widehat{\mathcal{R}}_{2,5}$, $\widehat{\mathcal{R}}^1_{2,5}$ or $\widehat{\mathcal{R}}^2_{2,5}$ in place under addition. From the proof of the metric regularity of the code $\mathcal{R}_{m-3,m}$ for odd m it is not hard to see that only the vectors from $\mathcal{R}_{2,5}$ do that, and thus the following holds:

Proposition 6.2 *Let $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}}+\tilde{\mathbf{v}}) \in \widehat{\widehat{\mathcal{R}}}_{2,6}$. Then $\tilde{\mathbf{u}} \in \mathcal{R}_{2,5}$.*

Recall from Section 3 that $\widehat{\mathcal{R}}_{1,5}$ is composed of 4 EA-equivalence classes: $\widehat{\mathcal{R}}_{1,5} = \bigcup_{i=1}^4 \widehat{\mathcal{R}}^i_{1,5}$. Somewhat similar to Lemma 6.1, the following statement holds:

Lemma 6.3 *For each $i = 1, 2, 3, 4$ one of the following statements holds:*

1. $\forall \mathbf{w}' \in \widehat{\mathcal{R}}^i_{1,5}\ \forall (\mathbf{y}, \mathbf{w}) \in \widehat{\mathcal{R}}_{2,6}\ \forall \mathbf{u} \in \mathcal{R}_{2,5}\ (d(\mathbf{y}, \mathbf{u}) = 6 \to \mathbf{w} + \mathbf{u} \neq \mathbf{w}')$;
2. $\forall \mathbf{w}' \in \widehat{\mathcal{R}}^i_{1,5}\ \exists (\mathbf{y}, \mathbf{w}) \in \widehat{\mathcal{R}}_{2,6}\ \exists \mathbf{u} \in \mathcal{R}_{2,5} : (d(\mathbf{y}, \mathbf{u}) = 6 \wedge \mathbf{w} + \mathbf{u} = \mathbf{w}')$;

The following result shows that any of the EA-equivalence classes of the metric complement of $\mathcal{R}_{1,5}$ are also rather "unstable" when summed with a non-affine function:

Lemma 6.4 *For any $\mathbf{v} \notin \mathcal{R}_{1,5}$ and any $i = 1, 2, 3, 4$ there exists a vector $\mathbf{w} \in \widehat{\mathcal{R}}^i_{1,5}$ such that $\mathbf{v} + \mathbf{w} \notin \widehat{\mathcal{R}}_{1,5}$.*

These last two lemmas allow us to show that for any $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}} + \tilde{\mathbf{v}}) \in \widehat{\widehat{\mathcal{R}}}_{2,6}$, the vector $\tilde{\mathbf{v}}$ is in $\mathcal{R}_{1,5}$. Combined with Proposition 6.2, this results in the

Theorem 6.5 *Let* $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}} + \tilde{\mathbf{v}}) \in \widehat{\widehat{\mathcal{R}}}_{2,6}$. *Then* $(\tilde{\mathbf{u}}, \tilde{\mathbf{u}} + \tilde{\mathbf{v}}) \in \mathcal{R}_{2,6}$.

Since the inverse inclusion holds for any linear code, Theorem 6.5 establishes the metric regularity of the code $\mathcal{R}_{2,6}$.

7 Conclusion

We have established the metric regularity of the codes $\mathcal{RM}(1,5)$, $\mathcal{RM}(2,6)$ and of the codes $\mathcal{RM}(k,m)$ for $k \geqslant m-3$. Factoring in the result by Tokareva [14], which proves the metric regularity of $\mathcal{RM}(1,m)$ for even m, we have covered all infinite families of Reed-Muller codes with known covering radius. The only other Reed-Muller codes with known covering radius, metric regularity of which has not been yet established, are $\mathcal{RM}(1,7)$ and $\mathcal{RM}(2,7)$. Given these results, we formulate the following

Conjecture 1 *All Reed-Muller codes* $\mathcal{RM}(k,m)$ *are metrically regular.*

References

[1] Berlekamp E., Welch L. *Weight distributions of the cosets of the* $(32,6)$ *Reed-Muller code.* IEEE Transactions on Information Theory. **18**(1), 203–207 (1972).

[2] Cohen, G., Honkala, I., Litsyn, S., Lobstein, A. *Covering codes.* Elsevier. **54**, (1997).

[3] Hou X. D. *Radius of the Reed-Muller code* $R(1,7)$ *– A Simpler Proof.* Journal of Combinatorial Theory, Series A. **74**(2), 337–341 (1996).

[4] Kutsenko, A. *Metrical properties of self-dual bent functions. Designs, Codes and Cryptography (2019).* doi:10.1007/s10623-019-00678-x

[5] McLoughlin A. M. *Covering Radius of the* $(m-3)$*-rd Order Reed Muller Codes and a Lower Bound on the* $(m-4)$*-th Order Reed Muller Codes.* SIAM Journal on Applied Mathematics. **37**(2), 419–422 (1979).

[6] Mesnager S.: *Bent Functions: Fundamentals and Results.* Springer International Publishing, (2016).

[7] Mykkeltveit J. *The covering radius of the* $(128,8)$ *Reed-Muller code is* 56. IEEE Transactions on Information Theory. **26**(3), 359–362 (1980).

[8] Oblaukhov A. K. *Metric complements to subspaces in the Boolean cube.* Journal of Applied and Industrial Mathematics. **10**(3), 397–403 (2016).

[9] Oblaukhov A. K. *Maximal metrically regular sets.* Siberian Electronic Mathematical Reports. **15**, 1842–1849 (2018).

[10] Oblaukhov A. *lower bound on the size of the largest metrically regular subset of the Boolean cube.* Cryptography and Communications. **11**(4), 777–791 (2019).

[11] Rothaus O. S. *On "bent" functions.* Journal of Combinatorial Theory, Series A. **20**(3), 300–305 (1976).

[12] Schatz J. *The second order Reed-Muller code of length* 64 *has covering radius* 18. IEEE Transactions on Information Theory. **27**(4), 529–530 (1981).

[13] Stanica P., Sasao T., Butler J. T. *Distance duality on some classes of Boolean functions.* Journal of Combinatorial Mathematics and Combinatorial Computing. 2018.

[14] Tokareva N. *Duality between bent functions and affine functions.* Discrete Mathematics. **312**(3), 666–670 (2012).

[15] Tokareva N. *Bent functions: results and applications to cryptography.* Academic Press, (2015).

[16] Wang Q. *The covering radius of the Reed–Muller code* $RM(2,7)$ *is* 40. Discrete Mathematics. **342**(12), Article 111625 (2019).

УДК 519.7 DOI 10.17223/2226308X/X/1

ON THE NUMBER OF UNSUITABLE BOOLEAN FUNCTIONS IN CONSTRUCTIONS OF FILTER AND COMBINING MODELS OF STREAM CIPHERS

T. A. Bonich, M. A. Panferov, N. N. Tokareva

Bonich T. A., Panferov M. A., Tokareva N. N. **ON THE NUMBER OF UNSUITABLE BOOLEAN FUNCTIONS IN CONSTRUCTIONS OF FILTER AND COMBINING MODELS OF STREAM CIPHERS.** It is well known that every stream cipher is based on a good pseudorandom generator. For cryptographic purposes we are interested in generation of pseudorandom sequences of the maximal possible period. A feedback register is one of the most known cryptographic primitives that is used in construction of stream generators. In this paper we analyze periodic properties of pseudorandom sequences produced by filter and combiner generators equipped with nonlinear Boolean functions. We determine which nonlinear functions in these schemes lead to pseudorandom sequences of not maximal possible period. We call such functions unsuitable and count the exact number of them for an arbitrary n.

Keywords: *stream cipher, filter generator, combiner generator, gamma, Boolean function*

Remember that a *feedback shift register (FSR)* contains two parts: a binary block $x = (x_{n-1}, \dots x_0)$ of length n and a feedback function $f : (x_{n-1}, \dots, x_0) \to \{0, 1\}$, where f is a Boolean function in n variables. First, we fill the block x with concrete values of bits; together they form the *initial state* of the register. For functioning of the FSR the time is considered to be discrete, i. e. it is divided into clock cycles. On each clock cycle, the value of $f(x)$ is calculated first, then the state $x = (x_{n-1}, \dots, x_1, x_0)$ of the register will be changed to the state $x' = (x_{n-2}, \dots x_0, f(x))$ while the bit x_{n-1} will be written as the first bit of the generated sequence *gamma*.

The properties of gamma generated by FSR are well studied in the case when f is a linear function. If f is nonlinear,[1], then there are too many open questions with properties of gamma that all are connected to analysis of nonlinear recurrent sequences, [2] and [3]. That is why in cryptography some nonlinear *combinations* of linear FSRs are considered, for instance, filter and combining models of stream generators based on LFSR, [4] and [5].

In this paper we analyze pseudorandom sequences produced by filter and combiner generators. Namely, we study which nonlinear functions h in these schemes lead to pseudorandom sequences such that their periods are not maximally possible. We call such functions *unsuitable* and count the exact number of them for an arbitrary n.

A *linear feedback shift register (LFSR)* consists of two parts: a binary vector $x = (x_{n-1}, \dots x_0)$ of length n and a linear feedback function f in n variables. A *state* of the register is a filling of vector x. During the encryption the register changes its states under an action of the feedback function. *Gamma* is a pseudorandom sequence generated by LFSR.

Also, LFSR can be specified using feedback polynomials. It is a polynomial of degree n defining bits to be summed. If $f(x_{n-1}, \dots, x_0) = a_0 x_{n-1} \oplus a_1 x_{n-2} \oplus \dots \oplus a_{n-1} x_0$, where $\oplus$ is sum modulo 2, then the corresponding feedback polynomial is defined as $p(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + 1$. If $p(z)$ is a primitive polynomial, then the period of a

[0]The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

pseudorandom sequence generated by LFSR is maximal, i.e. is equal to $2^n - 1$. Thererfore, linear feedback shift registers are usually considers with primitive polinomials.

0.1. Functions for the filter model

The filter generator consists of a single shift register of length n with a linear feedback and uses a primitive polynomial to change states. A Boolean function $h(x_{n-1}, \dots, x_0)$ applied to the current state generates a pseudorandom sequence gamma.

Let $\gamma = (y_1 y_2 \dots y_{2^n-1})$, where $y_1 = h(x_{n-1}, \dots, x_0), y_2 = h(x_{n-2}, \dots, x_0, f(x_{n-1}, \dots, x_0))$, etc. Since the number of all nonzero states is equal to $2^n - 1$, the maximal period of gamma is $2^n - 1$ too. In this paper we would like do determine all Boolean function h in n variables that lead to gammas with non-maximum period. Let us call such functions *unsuitable*.

Note that the number of them does not depend on a linear feedback function. But whether function is suitable or not for the given generator — it depends on the feedback function. When we count the number of unsuitable functions h, we do not consider a specific set of states. We say that there is a certain number of different states which the generator uses (all sets, that primitive polynomials generate, fit this definition). Next, we study which pseudorandom sequences will have the maximum length. We analyze the number of unsuitable sequences and then the number of unsuitable functions. Thus, our reasonings do not affect the specific order of the states. Accordingly, for any set of states which the generator uses, there will be the number of unsuitable functions h exactly that we calculated.

Theorem 1. Let n be an integer and $2^n - 1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s}$, where p_i are distinct prime numbers, α_i are positive integers, s is a some number. Then the number of unsuitable Boolean functions in n variables for the filter generator with LFSR based on a primitive polynomial is equal to

$$2 \sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$, and $+$ is a usual summing.

0.2. Functions for the combining model

Combiner generators use several linear feedback shift registers. Each register has its own length n_i, uses its primitive polynomial for changing states. A Boolean function $h(X_1, \dots, X_m)$ generates the pseudorandom sequence gamma where X_i is a register bit string i. Since we do not use the zero state in combiner generator the total number of states does not exceed $(2^{n_1} - 1)(2^{n_2} - 1) \dots (2^{n_m} - 1)$. In this case, the maximum is reached at $gcd(n_i, n_j) = 1$ where $i, j = 1, \dots, m$, $i \neq j$ and if all LFSRs have primitive feedback polynomials. Then the Boolean function can generate a gamma with period from 1 to $(2^{n_1} - 1)(2^{n_2} - 1) \dots (2^{n_m} - 1)$. Boolean functions h in n variables leading to gammas of non-maximum period in this case are called *unsuitable*.

We consider a more general model of a combiner generator. This generalized combining model is applied in ciphers such as Grain[6] and Bean[7]. Note that the classical combining model does not allow to describe a number of modern stream ciphers based on the more complicated operating with bits from different registers. In this case, the more known version of the combiner generator in which the function depends only on the extreme bits of the registers is included in the model we are considering. In a nonlinear model sometimes it is more convenient to work with several smaller registers than with one large register. It should be noted that the model that we consider can be used not only in cases of all linear or all non-linear registers but also in cases of mixed registers (i.e. some registers are linear, some are non-linear).

Theorem 2. Let n be an integer, $\sum_{i=1}^{m} n_i = n$. And

$$(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1) = p_1^{\alpha_1}p_2^{\alpha_2}\dots p_s^{\alpha_s},$$

where p_i are different prime numbers, $\alpha_i > 0$, s is an integer. Then the number of unsuitable Boolean functions in n variables for the combiner generator with LFSRs of lengths $n_1, \dots, n_m$ all based on primitive polynomials is equal to

$$2^{2^{n_1+n_2+\dots+n_m}-(2^{n_1}-1)(2^{n_2}-1)\dots(2^{n_m}-1)} \sum_{\beta\in\mathbb{F}_2^s,\beta\neq 0} ((-1)^{\beta_1+\dots+\beta_s+1} 2^{p_1^{\alpha_1-\beta_1}\dots p_s^{\alpha_s-\beta_s}}),$$

where $\beta = (\beta_1, \dots, \beta_s)$, and $+$ denotes a usual summing.

0.3. Functions for models with nonlinear registers

A *nonlinear feedback shift register (NFSR)* consists of two parts: a binary vector $x = (x_{n-1}, \dots x_0)$ of length n and a nonlinear state function $f : (x_{n-1}, \dots, x_0) \to \{0, 1\}$, in n variables.

Similarly to the linear case, consider the filter generator. We assume that NFSR passes over all 2^n states, i.e. it has maximal possible period.

Theorem 3. Let n be an integer. Then the number of unsuitable Boolean functions in n variables for the filter generator with NFSR of the maximal possible period is equal to $2^{2^{n-1}}$.

There is an another question related to NFSRs: how to determine for which nonlinear feedback functions NFSR of length n has the maximal possible period 2^n? This question is hard and still open.

We kindly thank the reviewer for careful reading of our paper and significant remarks.

ЛИТЕРАТУРА

1. *Key E.* An analysis of the structure and complexity of nonlinear binary sequence generators // IEEE Trans Inform Theory. 1976. № 22. p. 732–736.
2. *Gluhov M. M.,Elizarov V. P.,Nechaev A. A.* Algebra. Gelios ARV, 2003.
3. *Roman'kov V. A.* Introduction in cryptography. Lecture course, Forum, Moscow, 2012.
4. *Tokareva N. N.* Symmetric cryptography: a short course. Novosibirsk State University,2012.
5. *Carlet C.* Boolean functions for cryptography and error-correcting codes // Boolean Models and Methods in Mathematics, Computer Science, and Engineering / Eds. P. Hammer, Y. Crama. Cambridge Univ. Press, 2010. Chapter 8. p. 257—397. URL: www.math.univ-paris13.fr/∼carlet/
6. *Hell M., Johansson T., Meier W.* A Stream Cipher for Constrained Environments // Int. J. Wireless Mobile Comput., vol. 2, no. 1, 2007, p. 86–93.
7. *Kumar N., Ojha S., Jain K., Lal S.* BEAN: A lightweight stream cipher // Proceedings of the 2nd International Conference on Security of Information and Networks, SIN 2009. ACM, 2009, p. 168—171.

БОНИЧ Татьяна Андреевна — магистрантка Новосибирского государственного университета, исследователь в лаборатории криптографии JetBrains Research, г. Новосибирск. E-mail: **t.bonich@g.nsu.ru**

ПАНФЕРОВ Матвей Андреевич — магистрант Новосибирского государственного университета, исследователь в лаборатории криптографии JetBrains Research, г. Новосибирск. E-mail: **m.panferov@g.nsu.ru**

ТОКАРЕВА Наталья Николаевна — доцент, кандидат физико-математических наук, старший научный сотрудник в Институте математики им. С. Л. Соболева СО РАН, Новосибирсий государственный университет, заведующая лабораторией криптографии JetBrains Research, г. Новосибирск. E-mail: **tokareva@math.nsc.ru**

УДК 519.7

ON A SECONDARY CONSTRUCTION OF QUADRATIC APN FUNCTIONS

Kalgin K. V., Idrisova V. A.[1]

Almost perfect nonlinear functions possess the optimal resistance to the differential cryptanalysis and are widely studied. Most known constructions of APN functions are obtained as functions over finite fields $\mathbb{F}_{2^n}$ and very little is known about combinatorial constructions in $\mathbb{F}_2^n$. In this work we consider how to obtain a quadratic APN function in $n+1$ variables from a given quadratic APN function in n variables using special restrictions on new terms.

Ключевые слова: *Vectorial Boolean function, APN function, quadratic function, secondary construction*

Let us recall some definitions. Let $\mathbb{F}_2^n$ be the n-dimensional vector space over $\mathbb{F}_2$. A function F from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$, where n and m are integers is called a *vectorial Boolean function*. If $m = 1$ such a function is called *Boolean*. Every vectorial Boolean function F can be represented as a set of m *coordinate functions* $F = (f_1, \ldots, f_m)$, where f_i is a Boolean function in n variables. Any vectorial function F can be represented uniquely in its *algebraic normal form (ANF)*:

$$F(x) = \sum_{I \in \mathcal{P}(N)} a_I (\prod_{i \in I} x_i),$$

where $\mathcal{P}(N)$ is a power set of $N = \{1, \ldots, n\}$ and $a_I \in \mathbb{F}_2^m$. The *algebraic degree* of a given function F is the degree of its ANF: $\deg(F) = \max\{|I| : a_I \neq 0, I \in \mathcal{P}(N)\}$. If algebraic degree of a function F is not more than 1 then F is called *affine*. If for an affine function F it holds $F(\mathbf{0}) = \mathbf{0}$ then F is called *linear*. If algebraic degree of a function F is equal to 2 then F is called *quadratic*. Two vectorial functions F and G are *extended affinely equivalent (EA-equivalent)* if $F = A_1 \circ G \circ A_2 + A$ where A_1, A_2 are affine permutations on $\mathbb{F}_2^n$ and A is an affine function. Let F be a vectorial Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$. For vectors $a, b \in \mathbb{F}_2^n$, where $a \neq 0$, consider the value

$$\delta(a, b) = \big|\{\ x \in \mathbb{F}_2^n \ \big|\ F(x + a) + F(x) = b\}\big|.$$

Denote by Δ_F the following value:

$$\Delta_F = \max_{a \neq \mathbf{0},\ b \in \mathbb{F}_2^n} \delta(a, b).$$

Then F is called *differentially Δ_F-uniform* function. The smaller the parameter Δ_F is the better the resistance of a cipher containing F as an S-box to differential cryptanalysis. For the vectorial functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ the minimal possible value of Δ_F is equal to 2. In this case the function F is called *almost perfect nonlinear (APN)*. This notion was introduced by K. Nyberg in [7]. APN fuctions draw attention of many researchers, but there is still a significant list (see, for example, surveys [3], [6] or [8]) of important open questions. We are especially interested how to find new constructions of APN functions in vectorspace $\mathbb{F}_2^n$, since almost all the known constructions of this class are found only as

[1]The work was carried out within the framework of the state contract of the Sobolev Institute of Mathematics (project no. 0314-2019-0017) and supported by Russian Foundation for Basic Research (projects no. 18-07-01394, 20-31-70043) and Laboratory of Cryptography JetBrains Research.

polynomials over the finite fields, and to the best of our knowledge, only a few approaches to such combinatorial constructions was proposed, for example, in [4] and [5].

Recall that two vectorial functions F and G are *extended affinely equivalent (EA-equivalent)* if $F = A_1 \circ G \circ A_2 + A$ where A_1, A_2 are affine permutations on $\mathbb{F}_2^n$ and A is an affine function. Since EA-equivalence preserves APNness, it is always possible to omit linear and constant terms in the algebraic normal form of a given APN function. Further we will consider quadratic vectorial Boolean functions that have only quadratic terms in their ANF. The following theorem gives a necessary condition on the ANF of a given APN function.

Теорема 1. [1] Let $F = (f_1, \ldots, f_n)$ be an APN function in n variables. Then every quadratic term $x_i x_j$, where $i \neq j$, appears at least in one coordinate function of F.

This property motivated us to suggest the following construction of quadratic APN functions. Let $G = (g_1, \ldots, g_n)$ be a quadratic APN-function in n variables. Consider vectorial function $F = (f_1, \ldots, f_n, f_{n+1})$ in $n+1$ variables such that:

$$\begin{aligned} f_1 &= g_1 + \sum_{i=1}^{n} \alpha_{1,i} x_i x_{n+1}; \\ &\ldots \\ f_n &= g_n + \sum_{i=1}^{n} \alpha_{n,i} x_i x_{n+1}; \\ f_{n+1} &= g_{n+1} + \sum_{i=1}^{n} \alpha_{n+1,i} x_i x_{n+1}, \end{aligned} \tag{1}$$

where $\alpha_{1,i} \ldots, \alpha_{n+1,i} \in \mathbb{F}_2$ for $i = 1, \ldots, n$ and $g_{n+1} = \sum_{1 \leqslant j < k \leqslant n} \beta_{j,k} x_j x_k$ for some fixed $\beta_{j,k} \in \mathbb{F}_2$. Note that if $\alpha_{1,i}, \ldots, \alpha_{n,i}$ are such that each term $x_i x_{n+1}$ appears at least in one of the coordinate functions $f_1, \ldots, f_n$, then the necessary condition of Theorem 1 is held for the constructed function F.

Each quadratic vectorial function G in n variables can be considered as a symmetric matrix $\mathcal{G} = (g_{ij})$, where each element $g_{ij} \in \mathbb{F}_2^n$ is a vector of coefficients corresponding to term $x_i x_j$ in the algebraic normal form of G and all diagonal elements g_{ii} are null. It is necessary to mention that these matrices are essentially the same as so-called QAM matrices that were used in [10] and [9] to construct and classify a lot of new quadratic APN functions over finite fields. Using these matrices the APN property can be formulated in the following way:

Утверждение 1. Let $\mathcal{G}$ be the matrix that corresponds to quadratic vectorial function G. Then function G is APN if and only if $x \cdot (\mathcal{G} \cdot a) \neq 0$ for all $x \neq a$, where $a, x \in \mathbb{F}_2^n$ and $a \neq 0$.

In terms of matrices the construction from (1) can be considered as an extension of a given $\mathcal{G}$ with an extra bit that represents g_{n+1} in every element and an extra pair of row and column that represents a set of new terms $x_i x_{n+1}$.

Consider a quadratic APN function G and the corresponding $n \times n$ matrix $\mathcal{G}$. Denote the vector of nonzero coefficients as $\alpha = (\alpha_1, \ldots, \alpha_n)$. Let us fix g_{n+1} and construct $(n+1) \times (n+1)$ matrix $\mathcal{F}$ by adding $(\alpha_1, \ldots, \alpha_n, 0)$ as the last column and the last row and adding new bit to every element according to the choice of g_{n+1}. Let us denote as $\mathcal{G}'$ the submatrix (f_{ij}) of $\mathcal{F}$, such that $i, j < n+1$. Let $\langle X \rangle$ denote the linear span of X and F be the quadratic vectorial that is corresponded with the constructed matrix $\mathcal{F}$

Теорема 2. A function F is APN if and only if $\alpha \cdot a'$ does not belong to $\langle \mathcal{G}' \cdot a' \rangle$ for all $a' \in \mathbb{F}_2^n$, $a' \neq \mathbf{0}$.

This theorem shows how to choose new coefficients $\alpha_{1,i} \ldots, \alpha_{n+1,i} \in \mathbb{F}_2$ in the construction from (1) such that an obtained function F is APN. When $n = 3, 4$ and 5 for APN functions that are EA classes representatives we obtained all the possible classes of quadratic APN functions for $4, 5$ and 6 variables from the classification [2] and large variety of classes for constructing from 6 to 7 variables.

ЛИТЕРАТУРА

1. T. Beth, C. Ding. On almost perfect nonlinear permutations. *Advances in Cryptology, EUROCRYPT'93, Lecture Notes in Computer Science*, vol. 765, pp. 65-76, 1993.
2. M. Brinkmann, G. Leander. On the classification of APN functions up to dimension five. *Des. Codes Cryptogr.*, vol. 49, Issue 1–3, pp. 273-288, 2008.
3. C. Carlet. Open Questions on Nonlinearity and on APN Functions. *Arithmetic of Finite Fields. WAIFI 2014. Lecture Notes in Computer Science*, vol. 9061, pp 83-107 (2015).
4. A. A. Gorodilova. Characterization of almost perfect nonlinear functions in terms of subfunctions, *Diskr. Mat.*, vol. 27(3), pp. 3-16, 2015; *Discrete Math. Appl.*, vol. 26(4), pp. 193-202, 2016.
5. V. A. Idrisova. On an algorithm generating 2-to-1 APN functions and its applications to "the big APN problem". *Cryptogr. Commun.* 11, 21–39, 2019.
6. M. M. Glukhov. O priblizhenii diskretnykh funktsiy lineynymi funktsiyami [On the approximation of discrete functions by linear functions]. *Matematicheskie Voprosy Kriptografii*, 2016, vol. 7, no. 4, pp. 29–50. (in Russian)
7. K. Nyberg. Differentially uniform mappings for cryptography. *Advances in Cryptography, EUROCRYPT'93, Lecture Notes in Computer Science*, vol. 765, pp. 55-64, 1994.
8. M. E. Tuzhilin. APN-functions. *Prikl. Diskr. Mat.,* 2009, no. 3(5), 14–20
9. Y. Yu, N. S. Kaleyski, L. Budaghyan, Y. Li. Classification of quadratic APN functions with coefficients in GF(2) for dimensions up to 9. *IACR Cryptol. ePrint Arch.*: 1491, 2019.
10. Y. Yu, M. Wang, Y. Li. A matrix approach for constructing quadratic APN functions. *Des. Codes Cryptogr.* 73, 587-600, 2014

Kalgin K. V., Idrisova V. A. **ON A SECONDARY CONSTRUCTION OF QUADRATIC APN FUNCTIONS.** Almost perfect nonlinear functions possess the optimal resistance to the differential cryptanalysis and are widely studied. Most known constructions of APN functions are obtained as functions over finite fields $\mathbb{F}_{2^n}$ and very little is known about combinatorial constructions in $\mathbb{F}_2^n$. In this work we consider how to obtain a quadratic APN function in $n+1$ variables from a given quadratic APN function in n variables using special restrictions on new terms.
Keywords: *Vectorial Boolean function, APN function, quadratic function, secondary construction*

КАЛГИН Константин Викторович — младший научный сотрудник Института Математики им. С. Л. Соболева СО РАН и ассистент кафедры параллельных вычислений ФИТ Новосибирского Государственного Университета, г. Новосибирск. E-mail: **kalginkv@gmail.com**

ИДРИСОВА Валерия Александровна — научный сотрудник Института Математики им. С. Л. Соболева СО РАН, г. Новосибирск. E-mail: **vvitkup@yandex.ru**

UDC 519.7

ON ONE-TO-ONE PROPERTY OF A VECTORIAL BOOLEAN FUNCTION OF THE SPECIAL TYPE

M. M. Zapolskiy, N. N. Tokareva

Novosibirsk State University,Novosibirsk,Russia

E-mail: m.zapolskii@g.nsu.ru, tokareva@math.nsc.ru

S-boxes are widely used in cryptography. In particular, they form important components of SP and Feistel networks. Mathematically, S-box is a vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ that should satisfy several cryptographic properties. Usually $n = m$. In this work we study one-to-one property of a vectorial Boolean function constructed in a special way on the base of a Boolean function and a permuation on n elements. We count the number of all one-to-one functions of this type.

Keywords: *Boolean function, vectorial Boolean function, S-box.*

Let $\pi \in S_n$ be a permutation such that $\pi^n(x) = x$. Consider some $x \in \mathbb{F}_2^n$, $x = (x_1, \ldots, x_n)$, define $\pi(x) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. Let f be a Boolean function in n variables, we construct vectorial Boolean function $F_\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ by the rule, already mentioned in [1]:

$$F_\pi(x) = (f(x), f(\pi(x)), f(\pi^2(x)), \ldots, f(\pi^{n-1}(x))).$$

Let $\Delta_{\pi,n}$ be a set of these functions. Define $\rho(x) = (x_n, x_1, x_2, \ldots, x_{n-1})$, i.e. $\rho = (n, 1, 2, \ldots, n-1)$.

Proposition 1. Let $\pi \in S_n$ such that $\pi^n(x) = x$, $F_\pi \in \Delta_{\pi,n}$. Then $F_\pi(\pi(x)) = \rho^{-1}(F_\pi(x))$ for all $x \in \mathbb{F}_2^n$.

Further we consider that $\pi^n(x) = x$ for all $x \in \mathbb{F}_2^n$. Let π be an arbitrary permutation, we define action of π on $\mathbb{F}_2^n$ by the rule: if $x \in \mathbb{F}_2^n$ then $x \circ \pi = \pi(x)$. This action splits $\mathbb{F}_2^n$ into orbits with respect to π . If x is in some orbit o, we call x *a generator of* o. We call $O_\pi(x)$ *the orbit with respect to the action of* π.

Example: Let us give some examples of the orbits.

For $n = 4$ the set $\mathbb{F}_2^n$ is divided into six orbits with respect to the permutation ρ:

$O_\rho((0,0,0,0))$	$(0,0,0,0)$
$O_\rho((1,0,0,0))$	$(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$
$O_\rho((1,0,1,0))$	$(1,0,1,0), (0,1,0,1)$
$O_\rho((1,0,0,1))$	$(1,0,0,1), (1,1,0,0), (0,1,1,0), (0,0,1,1)$
$O_\rho((0,1,1,1))$	$(0,1,1,1), (1,0,1,1), (1,1,0,1), (1,1,1,0)$
$O_\rho((1,1,1,1))$	$(1,1,1,1)$

We denote by $\Theta_{\pi,n}$ a set of all orbits with repect to the action of π on $\mathbb{F}_2^n$. Proposition 1 implies that for arbitrary $F_\pi \in \Delta_{\pi,n}$ values of elements of some π-orbit $g \in \Theta_{\pi,n}$ are elements of some ρ-orbit $q \in \Theta_{\rho,n}$, since $F_\pi(\pi^i(x)) = \rho^{-i}(F_\pi(x))$. Let $M_{\pi,n}^k = \{g \in \Theta_{\pi,n} : |g| = k\}$.

Let $\Psi_{F_\pi,n} : \Theta_{\pi,n} \to \Theta_{\rho,n}$ be a mapping defined by the rule: $\Psi_{F_\pi,n}(O_\pi(x)) = O_\rho(F_\pi(x))$. Now we can formulate conditions for F_π to be one-to-one in terms of $\Psi_{F_\pi,n}$.

[0]Работа выполнена в рамках государственного задания ИМ СО РАН (проект № 0314-2019-0017) при поддержке Российского Фонда Фундаментальных Исследований (проект 18-07-01394) и лаборатории криптографии JetBrains Research

Theorem 1. $F_\pi \in \Delta_{\pi,n}$ is an one-to-one function if and only if $\Psi_{F_\pi,n}$ is one-to-one. If $\Psi_{F_\pi,n}$ is one-to-one, then $|\Psi_{F_\pi,n}(g)| = |g|$, for all $g \in \Theta_{\pi,n}$.

As a consequence of Theorem 1 we get the following result.

Proposition 2. If for some k it holds $|M^k_{\pi,n}| \neq |M^k_{\rho,n}|$ then the set of one-to-one functions from $\Delta_{\pi,n}$ is empty.

Theorem 1 means that in order to construct one-to-one functions $F_\pi \in \Delta_{\pi,n}$ we can use bijective maps $\Psi_n : \Theta_{\pi,n} \to \Theta_{\rho,n}$ that satisfy $|\Psi_n(g)| = |g|$, where $g \in \Theta_{\pi,n}$. Then, depending on them, we can construct $F_\pi \in \Delta_{\pi,n}$ such that $\Psi_{F_\pi,n} \equiv \Psi_n$.

Proposition 3. Let $\Psi_n : \Theta_{\pi,n} \to \Theta_{\rho,n}$ satisfy $|\Psi_n(g)| = |g|$ for all $g \in \Theta_{\pi,n}$. Then for all $k \in \mathbb{N}$ the restriction of Ψ_n on $M^k_{\pi,n}$ is a permutation of $M^k_{\pi,n}$.

Now consider the case $\pi = \rho$. We define $M^k_n = M^k_{\rho,n}$. Consider an one-to-one function Ψ_n which satisfies $|\Psi_n(g)| = |g|$ for all $g \in \Theta_{\pi,n}$. Let us construct function $F_\rho \in \Delta_{\rho,n}$ based on Ψ_n. Let $O \in \Theta_{\rho,n}$ be an orbit of length k. If value of F_ρ for some $x \in O$ is determined then value of F_ρ is determined for all $x \in O$, since $F_\rho(\rho^n(x)) = \rho^{-n}(F_\rho(x))$. Thus for every $\Psi_{F_\rho,n}$ we are able to construct $\prod_{k\in I_n} k^{|M^k_n|}$ functions, where $I_n = \{z \in \mathbb{N} : z|n\}$, and all of them are pairwise different.

Proposition 4. For any $k \in \mathbb{N}$ it holds $\sum_{\ell\in I_k} \ell \cdot |M^\ell_n| = 2^k$.

This formula allows us to calculate $|M^k_n|$ for every k. There are always only two orbits of length one, so we can calculate $|M^k_n|$ for every prime k. Then we can calculate it for every k. Therefore we get the number of one-to-one functions from $\Delta_{\rho,n}$ via the following result:

Theorem 2. The number of one-to-one vectorial Boolean functions in class $\Delta_{\rho,n}$ is equal to $\prod_{k\in I_n} |M^k_n|! \cdot k^{|M^k_n|}$.

REFERENCES

1. D. A. Zyubina, N. N. Tokareva. Cryptographic properties of a simple S-box construction based on a Boolean function and a permutation // Applied Discrete Mathematics. Supplement 2020.

Запольский Максим Михайлович — Новосибирский Государственный Университет. E-mail: **m.zapolskii@g.nsu.ru**

Токарева Наталья Николаевна — к.ф.-м.н.,Институт математики им. С. Л. Соболева, Новосибирск. E-mail: **tokareva@math.nsc.ru**

УДК 519.7

CRYPTOGRAPHIC PROPERTIES OF A SIMPLE S-BOX CONSTRUCTION BASED ON A BOOLEAN FUNCTION AND A PERMUTATION [1]

D. A. Zyubina, N. N. Tokareva

Sobolev Institute of Mathematics, Novosibirsk State University, Laboratory of Cryptography JetBrains Research, Novosibirsk, Russia

E-mail: d.zyubina@g.nsu.ru, tokareva@math.nsc.ru

We study a simple method of constructing S-boxes using Boolean functions and permutations. Let π be an arbitrary permutation on n elements, f be a Boolean function in n variables. Define a vectorial Boolean function $F_\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ as $F_\pi(x) = (f(x), f(\pi(x)), f(\pi^2(x)), \ldots, f(\pi^{n-1}(x)))$. In this paper we study cryptographic properties of F_π such as high nonlinearity, balancedness, low differential δ-uniformity in dependence on properties of f and π for small n.

Ключевые слова: *Boolean function, vectorial Boolean function, S-box, high nonlinearity, balancedness, low differential δ-uniformity, high algebraic degree.*

S-boxes play the crucial role for providing resistance of a block cipher to different types of attacks. The major reason for this that in classical and modern block ciphers the main complicated and nonlinear layer is presented namely by S-boxes. Mathematically, S-box is a vectorial Boolean function that maps n bits to m bits, or, $n \to m$. Usually, n coincides with m. It is well known that some special mathematical properties of S-boxes, such as high nonlinearity, low differential uniformity, high algebraic immunity, etc. make a cipher with such S-boxes be resistant to linear, differential, algebraic and other methods of cryptanalysis. It is well known that cryptographic properties of a Boolean (vectorial) function contradict to each other, [1], [2]. That is why we try to find vectorial Boolean functions that reach a *tradeoff* between different cryptographic properties and obligated to use mathematical methods (and not a direct computer search) for their constructing.

In this paper we propose a simple method of constructing S-boxes using Boolean functions. Let π be an arbitrary permutation on n elements, $\pi \in S_n$. If $x = (x_1, \ldots, x_n)$ is a binary vector then let $\pi(x)$ be a vector obtained as $\pi(x) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. Let f be a Boolean function in n variables. Define a vectorial Boolean function $F_\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ as follows

$$F_\pi(x) = (f(x), f(\pi(x)), f(\pi^2(x)), \ldots, f(\pi^{n-1}(x))).$$

In this paper we would like to study cryptographic properties of the vectorial Boolean function F_π in dependence on properties of the Boolean function f and the permutation π.

Note that this way of constructing vectorial Boolean functions was already mentioned before but only for obtaining some examples. Thus, A. Udovenko proposed a vectorial Boolean function of this type in 5 variables with the maximal possible algebraic immunity 3. It is a unique known solution of the previously unsolved problem from NSUCRYPTO 2016 [3]. So, functions F_π can have good crypto properties.

[1] The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2019-1613 with the Ministry of Science and Higher Education of the Russian Federation and Laboratory of Cryptography JetBrains Research.

Separately, we consider the special case of a permutation. Let A_n be the set of all full cycle permutations for n elements. For example, A_4 consists of 6 permutations: $(2,3,4,1)$, $(2,4,1,3)$, $(3,1,4,2)$, $(3,4,2,1)$, $(4,1,2,3)$, $(4,3,1,2)$ presented as vectors or (1234), (1243), (1342), (1324), (1432), (1423) in cyclic representation.

Let us recall definitions of several cryptographic properties.

A Boolean function f in n variables is called *balanced* if it takes every value (0 or 1) the same number of times [4]. A vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is *balanced* if it takes every value of $\mathbb{F}_2^n$ equally often [2] .

Let $\mathcal{A}_n = \{\langle a, x\rangle \oplus b : a \in \mathbb{F}_2^n, b \in \mathbb{F}_2\}$ be the class of all affine Boolean functions in n variables [5]. The *nonlinearity* $nl(f)$ of a Boolean function f in n variables is the Hamming distance between f and the set of all affine Boolean functions in n variables [5] . The *nonlinearity* $nl(F)$ of a vectorial Boolean function F is the minimal nonlinearity of all its component Boolean functions:

$$nl(F) = \min_{v\in\mathbb{F}_2^n} nl(F_v) = \min_{v\in\mathbb{F}_2^n} d(\langle v, F\rangle, \mathcal{A}_n) = \min_{v\in\mathbb{F}_2^n} \min_{g\in\mathcal{A}_n} d(\langle v, F\rangle, g),$$

where $v \neq 0$.

The *algebraic degree* of a vectorial Boolean function is the maximal algebraic degree of its component functions [2]. Note that for our construction $deg(F) = deg(f)$ for an arbitrary π, since all coordinate functions of F have degree $deg(f)$.

For a vectorial Boolean function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ let δ_F denote the maximal number of solutions for the equation $F(x) \oplus F(x \oplus a) = b$ while a, b run trough $\mathbb{F}_2^n$ and a is nonzero. Then F is called *differential δ_F-uniform*, see for instance [2]. Note that the minimal possible value of δ_F, where F maps from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ is 2.

We consider cryptographic properties of F_π for small n in relation to f and π.

All of the following propositions are obtained via computer search.

1. Case $n = 2$

• For any permutation $\pi \in S_2$ there exists a Boolean function f in 2 variables such that $\delta_{F_\pi} = 2$. Moreover, such Boolean functions are constructed as $f(x) = x_1x_2 \oplus a_1x_1 \oplus a_2x_2 \oplus a_0$, where $a_0, a_1, a_2 \in F_2$.

2. Case $n = 3$

For any Boolean function f in 3 variables $nl(f) \leqslant 2$.

• For any permutation $\pi \in A_3$ there exists a balanced Boolean function f in 3 variables such that vectorial Boolean function F_π is balanced.

• For any permutation $\pi \in A_3$ it holds $nl(F_\pi) = nl(f)$. Note that if $nl(F_\pi) = 2$, i.e. is maximal, then $\delta_{F_\pi} = 2$, i.e. is minimal possible. The number of such functions f is 48.

• For arbitary permutation $\pi \notin A_3$ and Boolean function f in 3 variables $\delta_{F_\pi} \geqslant 4$.

3. Case $n = 4$

Let us introduce the notation for permutations from the set A_4: $\pi_1 = (2,3,4,1)$, $\pi_2 = (4,1,2,3)$, $\pi_3 = (2,4,1,3)$, $\pi_4 = (3,1,4,2)$, $\pi_5 = (3,4,2,1)$, $\pi_6 = (4,3,1,2)$. Note that $\pi_1^{-1} = \pi_2, \pi_3^{-1} = \pi_4, \pi_5^{-1} = \pi_6$.

• For any permutation $\pi \in A_4^1$ and a balanced Boolean function f in 4 variables such that $\delta_{F_\pi} = 2$, F_π is not balanced.

• For any permutation $\pi \in A_4^1$ there exists a Boolean function f in 4 variables such that if $\delta_{F_\pi} = 2$ and nonlinearity of f and F_π are the same then $\delta_{F_{\pi^{-1}}} = 2$. Moreover, nonlinearity

of $F_{\pi^{-1}}$ and f coincide.

• For any permutation $\pi \notin A_4^1$ for arbitary Boolean function f in 4 variables $\delta_{F_\pi} \geqslant 4$.

Based on the results, we suppose that it is possible to construct vectorial Boolean functions in the arbitrary number of variables with cryptographic properties good enough using our simple construction for necessary Booleans functions and permutations.

We plan to use our program for studying vectorial Boolean functions with larger number of variables, now this work is in progress.

ЛИТЕРАТУРА

1. Cusick T. W., Stănică P. Cryptographic Boolean Functions and Applications. USA: Acad. Press. Elsevier, 2009.
2. Carlet C., "Vectorial Boolean Functions for Cryptography in Boolean Models and Methods in Mathematics, Computer Science, and Engineering, Crama Y. and Hammer P. L., Eds. Cambridge: Cambridge University Press, 2010, pp. 398-470.
3. Tokareva N., Gorodilova A., Agievich S., Idrisova V., Kolomeec N., Kutsenko A., Oblaukhov A., Shushuev G. Mathematical methods in solutions of the problems from the Third International Students? Olympiad in Cryptography. Prikladnaya Diskretnaya Matematika (Applied Discrete Mathematics). 2018, No. 40, pp. 34-58.
4. Carlet C., "Boolean Functions for Cryptography and Error-Correcting Codes in Boolean Models and Methods in Mathematics, Computer Science, and Engineering, Crama Y. and Hammer P. L., Eds. Cambridge: Cambridge University Press, 2010, pp. 257-397.
5. Logachev O.A., Salnikov A.A., Smyshlyaev S.V., Yaschenko V.V., "Boolean functions in coding theory and cryptology MCCME, 2012.

Zyubina D. A., Tokareva N. N. **CRYPTOGRAPHIC PROPERTIES OF A SIMPLE S-BOX CONSTRUCTION BASED ON A BOOLEAN FUNCTION AND A PERMUTATION.** We study a simple method of constructing S-boxes using Boolean functions and permutations. Let π be an arbitrary permutation on n elements, f be a Boolean function in n variables. Define a vectorial Boolean function $F_\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ as $F_\pi(x) = (f(x), f(\pi(x)), f(\pi^2(x)), \ldots, f(\pi^{n-1}(x)))$. In this paper we study cryptographic properties of F_π such as high nonlinearity, balancedness, low differential δ-uniformity in dependence on properties of f and π for small n.

Keywords: *Boolean function, vectorial Boolean function, S-box, high nonlinearity, balancedness, low differential δ-uniformity, high algebraic degree.*

ЗЮБИНА Дарья Александровна — м.н.с. Института математики им. С. Л. Соболева, студентка факультета информационных технологий НГУ, исследователь лаборатории криптографии JetBrains Research, Новосибирск. E-mail: **d.zyubina@g.nsu.ru**

ТОКАРЕВА Наталья Николаевна — к.ф.-м.н., с.н.с. Института математики им. С. Л. Соболева, доцент НГУ, зав. лаборатории криптографии JetBrains Research, Новосибирск. E-mail: **tokareva@math.nsc.ru**

УДК 519.7

О ДИФФЕРЕНЦИАЛАХ ДЛЯ МОДИФИКАЦИИ ШИФРА SIMON НА ОСНОВЕ СХЕМЫ ЛАЯ — МЭССИ[1]

А. А. Белоусова, Н. Н. Токарева

E-mail:

Рассматриваются блочный итеративный шифр Simon 32/64, основанный на сети Фейстеля, и его модификации на основе схемы Лая — Мэсси. Получены оценки вероятностей дифференциалов 12 раундов исходного шифра и его модификаций.

Ключевые слова: *схема Лая — Мэсси, сеть Фейстеля, дифференциальный криптоанализ.*

В работе рассматриваются блочные итеративные шифры, основанные на сети Фейстеля (рис. 1) и на альтернативной схеме — схеме Лая — Мэсси [1] (рис. 2). Для исследования выбран шифр Simon 32/64 [2], основанный на сети Фейстеля, и построены две его модификации подстановкой схемы Лая — Мэсси на место сети Фейстеля. Получены оценки для вероятностей дифференциалов, построенных для 12 раундов исходной и модифицированных версий шифра Simon 32/64. Оценка для вероятности дифференциалов для шифра Simon взята из работы [3], где получено, что для Simon 32/64 максимальная вероятность дифференциала после прохождения 12 раундов составляет 2^{-36}.

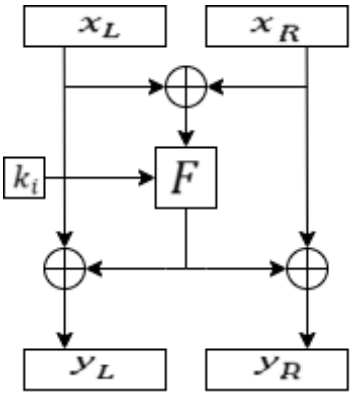


Рис. 1. Сеть Фейстеля

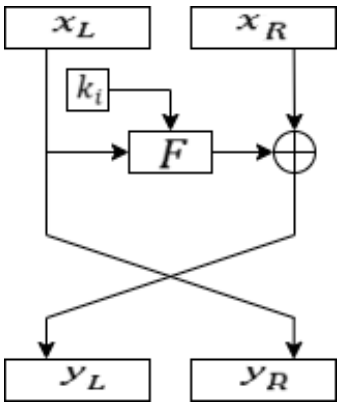


Рис. 2. Схема Лая-Мэсси

Один раунд схемы Лая — Мэсси в её оригинальном виде записывается как $(y_L, y_R) = (x_L \oplus F(x_L \oplus x_R), x_R \oplus F(x_L \oplus x_R))$, и в этом есть существенный недостаток: для любого входа (x_L, x_R) выполняется соотношение $x_L \oplus x_R = y_L \oplus y_R$, где (y_L, y_R) — выход раунда. В работе [4] отмечено, что для устранения этого недостатка к схеме необходимо добавить перестановку-ортоморфизм σ.

Пусть $\sigma\colon \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ — перестановка на $\mathbb{Z}_2^n$; σ называется *ортоморфизмом* $\mathbb{Z}_2^n$, если $\sigma \oplus I$ — также перестановка на $\mathbb{Z}_2^n$, где I — тождественная перестановка. Тогда один

[1]Работа выполнена в рамках государственного задания ИМ СО РАН (проект № 0314-2019-0017) при поддержке РФФИ (проект № 18-07-01394) и лаборатории криптографии JetBrains Research.

раунд схемы записывается как $(y_L, y_R) = (\sigma(x_L \oplus F(x_L \oplus x_R)), x_R \oplus F(x_L \oplus x_R))$, а разница текстов $y_L \oplus y_R = \sigma(x_L \oplus F(x_L \oplus x_R)) \oplus (x_R \oplus F(x_L \oplus x_R))$.

Проведено сравнение оценок вероятностей дифференциалов [5] оригинальной схемы Лая — Мэсси и схемы с добавлением ортоморфизма. Для этого написана программа, которая реализует перебор всех разностей открытых текстов. На каждой итерации шифра находится один из наиболее вероятных выходов на раунде с помощью построения строки таблицы дифференциалов, соответствующей входной разности. Далее найденные вероятности перемножаются для получения оценки максимальной вероятности дифференциалов.

Было получено, что после 12 раундов оценка максимальной вероятности дифференциала для модернизированного шифра Simon32/64 без добавления ортоморфизма составляет 2^{-24}, а с добавлением ортоморфизма находится в интервале между 2^{-24} и 2^{-63}.

Таким образом, оценка максимальной вероятности дифференциала модернизации шифра Simon 32/64 без добавления ортоморфизма выше, чем у оригинального шифра. Компьютерные вычисления на части данных позволяют предположить, что модернизация с ортоморфизмом может быть более устойчивой, чем оригинальный шифр и модернизация без ортоморфизма.

ЛИТЕРАТУРА

1. *Nakahara J.* Lai — Massey Cipher Designs. History, Design Criteria and Cryptanalysis. Springer Nature Switzerland AG, 2018.
2. *Beaulieu R., Shors D., Smith J., et al.*The Simon and Speck Families Of Lightweight Block Ciphers. Cryptology ePrint Archive, Report 2013/404, 2013.
3. *Abed F., List E., Lucks S., and Wenzel J.* Differential and Linear Cryptanalysis of Reduced-Round Simon. ePrint Archive, Report 2013/526, 2013.
4. *Vaudenay S* On the Lai — Massey Scheme // ASIACRYPT'99. LNCS. 1999. V. 1716. P. 8–19.
5. *Biham E. and Shamir A.* Differential Cryptanalysis of the Data Encryption Standard. Berlin; Heidelberg: Springer, 1993.

Belousova A. A., Tokareva N. N. **ON DIFFERENTIALS FOR THE MODIFICATION OF THE CIPHER SIMON BASED ON THE LAI — MESSI SCHEME.** We consider the block iterative cipher Simon based on the Feistel network and its modification based on the Lai -— Messi scheme. Received estimates of differentials of the considered ciphers are compared. The results show that after 12 rounds, estimate of the maximum probability of a differential for the modified cipher Simon 32/64 without adding an orthomorphism is 2^{-24}, and with the addition of orthomorphism is between 2^{-24} and 2^{-63}, while the estimate of maximum probability for the original version is 2^{-36}.

Keywords: *Lay — Massey scheme, Feistel network, differential cryptanalysis.*

БЕЛОУСОВА Алина Александровна — младший научный сотрудник Института математики им. С. Л. Соболева СО РАН, студентка Новосибирского государственного университета, г. Новосибирск. E-mail: **alinkabel18@gmail.com**

ТОКАРЕВА Наталья Николаевна — кандидат физико-математических наук, старший научный сотрудник Института математики им. С. Л. Соболева СО РАН, доцент Новосибирского государственного университета, г. Новосибирск. E-mail: **tokareva@math.nsc.ru**

УДК 519.7

ПОИСК КРИПТОГРАФИЧЕСКИХ БУЛЕВЫХ ФУНКЦИЙ С ПОМОЩЬЮ SAT-РЕШАТЕЛЕЙ [1]

А. Е. Доронин, К. В. Калгин

Новосибирский Государственный Университет,
лаборатория криптографии JetBrains Research, г. Новосибирск, Россия

E-mail: artem96dor@gmail.com, kalginkv@gmail.com

В данной работе представлен подход к решению некоторых криптографических задач, основанный на их сведении к классической задаче о выполнимости и на последующем использовании SAT-решателей. Представлено построение нескольких формул, определяющих условия взаимной однозначности и дифференциальной равномерности векторной булевой функции, а также проверяющих EA-эквивалентность двух булевых функций.

Ключевые слова: *SAT-решатели, криптография, булевы функции, EA-эквивалентность.*

В настоящее время SAT-решатели используются для решения криптографических задач разного типа. Например, для проведения криптоанализа асимметричной криптосистемы RSA [1], семейства шифров Trivium [2]. В [3] была представлена гомоморфная криптосистема с открытым ключом, основанная на SAT-задаче. Также с помощью SAT-решателей успешно проводилась проверка обратимости известных векторных булевых функций [4].

В данной работе предлагается использование SAT-решателей в задачах поиска криптографических булевых функций и проверки эквивалентности двух булевых функций. Для получения набора булевых формул были использованы следующие понятия и свойства.

Определение 1. Векторная булева функция $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ называется *взаимно однозначной*, если она инъективна или сюръективна, то есть выполняется одно из следующих условий:

1) $\forall x' \in \mathbb{Z}_2^n \; \forall x'' \in \mathbb{Z}_2^n : x' \neq x'' \to F(x') \neq F(x'')$,
2) $\forall y \in \mathbb{Z}_2^n \; \exists x \in \mathbb{Z}_2^n : F(x) = y$.

Определение 2. Векторная булева функция $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ является *дифференциально δ-равномерной*, если для любого ненулевого $a \in \mathbb{Z}_2^n$ и произвольного $b \in \mathbb{Z}_2^n$ уравнение $F(x) \oplus F(x \oplus a) = b$ имеет не более δ решений.

Определение 3. Векторные булевы функции F и G, действующие из $\mathbb{Z}_2^n$ в $\mathbb{Z}_2^n$, называются *EA-эквивалентными*, если выполняется: $G = B \circ F \circ A + C$, где A, B и C — аффинные функции.

Условия, фигурирующие в данных определениях, представляются в виде КНФ и подаются на вход SAT-решателя. В результате его работы происходит означивание переменных таким образом, чтобы формулы были истинными, а следовательно условия выполнялись.

[1]Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 18-07-01394) и лаборатории криптографии JetBrains Research.

Векторные булевы функции были закодированы в двух предсталениях при помощи соответствующих булевых переменных.

$f_{x,y} = 1 \iff F(x) = y$, где $x, y \in \mathbb{Z}_2^n$ — разреженное представление.

$fb_{x,k} = 1 \iff F_k(x) = 1$, где $k = 0, \ldots, n-1$, $x \in \mathbb{Z}_2^n$ — плотное представление.

Определения выше определяются следующими формулами.

Теорема 1. Множество переменных $f_{x,y}$ кодирует функцию F тогда и только тогда, когда следующая формула является истинной:

$$\mathbf{F^S}(f) = \bigwedge_{x \in \mathbb{Z}_2^n} \Big(\bigwedge_{\substack{y', y'' \in \mathbb{Z}_2^n \\ y' < y''}} \overline{f_{x,y'}} \vee \overline{f_{x,y''}} \Big) \wedge \bigwedge_{x \in \mathbb{Z}_2^n} \Big(\bigvee_{y \in \mathbb{Z}_2^n} f_{x,y} \Big). \tag{1}$$

Теорема 2. Множество переменных $f_{x,y}$ кодирует взаимно однозначную функцию тогда и только тогда, когда выполняются условия **Теоремы 1**, и следующая формула является истинной:

$$\mathbf{P^S}(f) = \bigwedge_{y \in \mathbb{Z}_2^n} \Big(\bigwedge_{\substack{x', x'' \in \mathbb{Z}_2^n \\ x' < x''}} \overline{f_{x',y}} \vee \overline{f_{x'',y}} \Big) \wedge \bigwedge_{y \in \mathbb{Z}_2^n} \Big(\bigvee_{x \in \mathbb{Z}_2^n} f_{x,y} \Big). \tag{2}$$

Теорема 3. Переменные $fbq_{x,y,k}$ и $fb_{x,k}$ кодируют взаимно однозначную векторную булеву функцию тогда и только тогда, когда следующая формула является истинной:

$$\mathbf{P^D_{sum}}(fb, fbq) = \bigwedge_{x,y \in \mathbb{Z}_2^n} \bigvee_{k} fbq_{x,y,k} \wedge \mathbf{SoP^D}(fb, fbq). \tag{3}$$

Теорема 4. Переменные $f_{x,y}$ и $fb_{x,k}$ кодируют взаимно однозначную векторную булеву функцию тогда и только тогда, когда следующая формула является истинной:

$$\mathbf{P^D_{sparse}}(f, fb) = \bigwedge_{x} \bigvee_{y \neq x} f_{x,y} \wedge \mathbf{SpDen}(f, fb). \tag{4}$$

Теорема 5. Отображение $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ является APN-функцией тогда и только тогда, когда выполняются условия Теоремы 1, и следующая формула является истинной:

$$\mathbf{APN^S}(f, d) = \mathbf{Der^S}(f, d) \wedge \bigwedge_{\substack{b \neq 0, a \neq 0, \\ x, y \neq x}} (\overline{d_{x,a,b}} \vee \overline{d_{y,a,b}}). \tag{5}$$

Теорема 6. Переменные $fb_{x,k}$, $fbq_{x,x\oplus a,k}$ и $dbq_{x,y,a,k}$ кодируют APN-функцию тогда и только тогда, когда следующая формула является истинной:

$$\mathbf{APN^D}(fb, fbq, dbq) = \mathbf{SoPEq^D}(fbq, dbq) \wedge \mathbf{SoP^D}(fb, fbq) \wedge \bigwedge_{a,x,y} \bigvee_{k} \overline{dbq_{x,y,a,k}}. \tag{6}$$

Теорема 7. Переменные $A_{i,j}$, $B_{i,j}$, $C_{i,j}$, a_i и c_i кодируют EA-эквивалентность булевых функций, если следующая формула является истинной:

$$\mathbf{EA}(fb, gb) = \bigwedge_{x} \Big(\mathbf{MatVec}(y, A, x, a) \wedge \mathbf{MatVec}(r, C, x, c) \wedge \mathbf{MatVec}(z, B, \mathbf{gb_y}, 0) \wedge \mathbf{P^D_{sum}}(y, yq) \wedge \mathbf{P^D_{sum}}(z, zq) \Big) \wedge \mathbf{SoP^D_{EA}}(fb, z, r),$$

где $\mathbf{MatVec}(y, A, x, c) = 1 \iff y = A \cdot x \oplus c$, $\mathbf{SoP}^{\mathbf{D}}_{\mathbf{EA}}(f, z, r) = \bigwedge\limits_{x,k}(fb_{x,k} \vee z_k \vee \overline{r_k}) \wedge (fb_{x,k} \vee \overline{z_k} \vee r_k) \wedge (\overline{fb_{x,k}} \vee z_k \vee r_k) \wedge (\overline{fb_{x,k}} \vee \overline{z_k} \vee \overline{r_k})$.

В данной работе представлен набор формул для поиска криптографических функций и проверки EA-эквивалентности двух булевых функций при помощи SAT-решателей. Было описано построение формул для взаимной однозначности, дифференциальной равномерности, а также EA-эквивалентности двух векторных булевых функций. Входной файл для SAT-решателя генерируется на основе этих формул. Полученный набор формул также можно использовать для тестирования работы новых SAT-решателей, созданных для криптографических задач.

ЛИТЕРАТУРА

1. *Огородников Ю. Ю.* Комбинированная атака на алгоритм RSA с использованием sat-подхода Динамика систем, механизмов и машин Омск: ОмГТУ, 2016. С. 276–284.
2. *Заикин О. С., Отпущенников И. В., Семёнов А. А.* Оценки стойкости шифров семейства Trivium к криптоанализу на основе алгоритмов решения проблемы булевой выполнимости // ПДМ. Приложение. 2016. № 9. С. 46–48.
3. *Schmittner S. E.* A SAT-based Public Key Cryptography Scheme // IACR Cryptol. ePrint Arch. 2015. № 2015. С. 771.
4. *Wille R., Lye A., Niemann P.* Checking Reversibility of Boolean Functions // Reversible Computation: 8th International Conference Italy: RC, 2016. С. 322–337.

Doronin A. E., Kalgin K. V. **CONSTRUCTION OF CRYPTOGRAPHIC BOOLEAN FUNCTIONS USING SAT-SOLVERS.** In this paper we propose a method of solving some cryptographic problems based on translation them into SAT-problems and application of SAT-solvers. We introduce construction of several formulas defining conditions of one-to-one property and differential uniformity of vectorial Boolean functions and formulas for checking extended-affine equivalence for two vectorial Boolean functions.
Keywords: *SAT-solvers, cryptography, Boolean functions, EA-equivalence.*

ДОРОНИН Артемий Евгеньевич — студент Новосибирского государственного университета, г.Новосибирск. E-mail: **artem96dor@gmail.com**

КАЛГИН Константин Викторович — к.ф.-м.н., н.с. Института вычислительной математики и математической геофизики СО РАН, старший преподаватель Новосибирского государственного университета, г.Новосибирск. E-mail: **kalginkv@gmail.com**

УДК 004.056.55

КРИПТОАНАЛИЗ БАЗОВОЙ ВЕРСИИ КРИПТОСИСТЕМЫ С ОТКРЫТЫМ КЛЮЧОМ, ОСНОВАННОЙ НА СЛОЖНОСТИ РЕШЕНИЯ СИСТЕМЫ ПОЛИНОМИАЛЬНЫХ УРАВНЕНИЙ В ЦЕЛЫХ ЧИСЛАХ[1]

Е. В. Завалишина

Новосибирский государственный университет
Институт математики им. С. Л. Соболева
Лаборатория криптографии JetBrains Research

E-mail: e.zavalishina@g.nsu.ru

В работе представлены сведения о криптоанализе базовой версии криптосистемы с открытым ключом, основанной на сложности решения систем полиномиальных уравнений в целых числах. Автором работы был разработан алгоритм атаки на основе подобранного открытого текста, позволяющий получить набор матриц, которые могут использоваться в качестве секретного ключа. Было обнаружено, что набор таких матриц не единственный, а также выявлены некоторые свойства этих наборов.

Ключевые слова: *открытый ключ, криптоанализ, постквантовая криптография, полиномиальные уравнения.*

В 2016 году the National Institute of Standards and Technology представил доклад под названием Report on Post-Quantum Cryptography [1], в котором полагает, что пришло время подготовиться к переходу на квантово-устойчивую криптографию, так как некоторые задачи, лежащие в основе использующихся на практике криптографических алгоритмов, могут быть решены квантовыми компьютерами.

В связи с этим автором настоящей работы и соавторами была предпринята попытка создать новый алгоритм шифрования данных с открытым ключом, основанный на решении системы однородных полиномиальных уравнений в целых числах [2]. Кратко опишем основной принцип работы.

Имеем открытый текст $P = (P_1, ..., P_n)^T$, где P_i — целое число от 0 до $2^b - 1$ для некоторого положительного b. Пусть $K_{priv} = \{m, A, B\}$, $K_{pub} = \{f(x)\}$, где

— $m = m_1 m_2 ... m_k$ — целочисленный модуль, такой, что $m > 2^b$, а $\varphi(m_i) \bmod 3 \neq 0$ для любого m_i;
— A и B — целочисленные матрицы $n \times n$;
— $f(x) = (f_1(x), ..., f_n(x))^T$, где $f_i(x)$ — полиномы от вектора переменных $x = (x_1, ..., x_n)^T$, вычисленные в три шага по модулю m:

1) $u(x) = A \times x$;
2) $s(x) = ((u_1(x))^3, ..., (u_n(x)^3))$;
3) $f(x) = B \times s(x)$.

Шифртекст $C = (C_1, ..., C_n)^T$ вычисляется как $C = f(P_1, ..., P_n)$.

Данная работа посвящена криптоанализу описанной системы.

[1]Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации номер 075-15-2019-1613 и лаборатории криптографии JetBrains Research.

Автором работы был разработан алгоритм атаки на основе подобранного открытого текста, который позволяет получить набор матриц — эквивалентных ключей системы, которые могут использоваться в качестве секретного ключа.

Набор полиномов, использующийся в качестве открытого ключа в базовой версии, имеет строго определенный вид. Рассмотрим общий вид системы полиномов на простейшем примере с двумя переменными.

Пусть матрицы A и B имеют следующий вид:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Тогда открытый ключ выглядит так:

$$\begin{aligned} &x_1^3(a_{11}^3 b_{11} + a_{21}^3 b_{12}) + x_1^2 x_2(3a_{11}^2 a_{12} b_{11} + 3a_{21}^2 a_{22} b_{12}) + \\ &\quad + x_1 x_2^2(3a_{11} a_{12}^2 b_{11} + 3a_{21} a_{22}^2 b_{12}) + x_2^3(a_{12}^3 b_{11} + a_{22}^3 b_{12}) = \\ &\quad = \alpha_1 x_1^3 + \alpha_2 x_1^2 x_2 + \alpha_3 x_1 x_2^2 + \alpha_4 x_2^3 \\ &x_1^3(a_{11}^3 b_{21} + a_{21}^3 b_{22}) + x_1^2 x_2(3a_{11}^2 a_{12} b_{21} + 3a_{21}^2 a_{22} b_{22}) + \\ &\quad + x_1 x_2^2(3a_{11} a_{12}^2 b_{21} + 3a_{21} a_{22}^2 b_{22}) + x_2^3(a_{12}^3 b_{21} + a_{22}^3 b_{22}) = \\ &\quad = \alpha_5 x_1^3 + \alpha_6 x_1^2 x_2 + \alpha_7 x_1 x_2^2 + \alpha_8 x_2^3. \end{aligned} \tag{1}$$

Для нахождения эквивалентных ключей необходимо выразить коэффициенты полиномов открытого ключа (1) через элементы матриц A и B и решить систему сравнений (2). Набор решений системы сравнений (2) является набором эквивалентных ключей криптосистемы. Имеем

$$\begin{aligned} a_{11}^3 b_{11} + a_{21}^3 b_{12} &\equiv \alpha_1 \pmod m \\ 3a_{11}^2 a_{12} b_{11} + 3a_{21}^2 a_{22} b_{12} &\equiv \alpha_2 \pmod m \\ 3a_{11} a_{12}^2 b_{11} + 3a_{21} a_{22}^2 b_{12} &\equiv \alpha_3 \pmod m \\ a_{12}^3 b_{11} + a_{22}^3 b_{12} &\equiv \alpha_4 \pmod m \\ a_{11}^3 b_{21} + a_{21}^3 b_{22} &\equiv \alpha_5 \pmod m \\ 3a_{11}^2 a_{12} b_{21} + 3a_{21}^2 a_{22} b_{22} &\equiv \alpha_6 \pmod m \\ 3a_{11} a_{12}^2 b_{21} + 3a_{21} a_{22}^2 b_{22} &\equiv \alpha_7 \pmod m \\ a_{12}^3 b_{21} + a_{22}^3 b_{22} &\equiv \alpha_8 \pmod m. \end{aligned} \tag{2}$$

Уравнения такого вида можно построить для любых n и m, удовлетворяющих условиям для параметров криптосистемы.

Так как система сравнений нелинейна и не существует алгоритма, позволяющего эффективно решать такие системы, решение производится перебором.

Между матрицами также были отмечены линейные зависимости.

Утверждение 1. Для описанной криптоистемы с любыми n и m, удовлетворяющими условиям, можно выразить коэффициенты полиномов открытого ключа через элементы матриц секретного ключа и получить систему сравнений по модулю m. Любое решение этой системы сравнений является эквивалентным секретным ключом криптосистемы.

Утверждение 2. Для любого $K_{priv} = \{A, B, m\}$, удовлетворяющего условиям для параметров криптосистемы, существуют эквивалентные ключи, образованные всеми возможными перестановками строк матрицы A и соответствующих столбцов матрицы B. Утверждение верно для любых n.

В таблице 1 указаны некоторые экспериментальные данные. Обозначим количество эквивалентных ключей K_{eq}, а количество всех возможных наборов матриц A и B для количества переменных n и модуля m — M_{total}.

Таблица 1

Количество эквивалентных ключей

n	m	K_{eq}	M_{total}	K_{eq}/M_{total}
2	11	200	214 358 881	$9{,}3\text{e}^{-7}$
2	17	512	6 975 757 441	$7{,}3\text{e}^{-8}$
2	23	968	78 310 985 281	$1{,}2\text{e}^{-8}$
2	29	1568	500 246 412 961	$3{,}1\text{e}^{-9}$
2	41	3200	7 984 925 229 121	4e^{-10}
2	53	5408	62 259 690 411 361	$0{,}1\text{e}^{-10}$
2	59	6728	146 830 437 604 321	$4{,}5\text{e}^{-11}$
3	11	6000	5 559 917 313 492 231 481	$0{,}1\text{e}^{-15}$

Из экспериментальных данных можно предположить, что криптографическая стойкость рассматриваемой системы растет с увеличением модуля и количества переменных.

ЛИТЕРАТУРА

1. *National Institute of Standards and Technology* NIST Internal or Interagency Reports (IR) 8105 Report on Post-Quantum Cryptography, Gaithersburg, Maryland, April 2016, 15pp.
2. *Волков, Е., Баранов А., Завалишина Е.* Криптографическая система с открытым ключом Second Conference on Software Engineering and Information Management (SEIM-2017), 2017. — С. 41-44.

Zavalishina E. V. **CRYPTANALYSIS OF THE BASIC VERSION OF A PUBLIC KEY CRYPTOSYSTEM BASED ON THE COMPLEXITY OF SOLVING A SYSTEM OF POLYNOMIAL EQUATIONS WITH AN INTEGER SOLUTION.** The paper provides information about cryptanalysis of the basic version of a public key cryptosystem based on the complexity of solving a system of polynomial equations with an integer solution. The author of the theses developed the chosen-plaintext attack algorithm, which allows to obtain a set of matrices that can be used as a secret key.

It was found that the set of such matrices is not the only one, and also some properties of these sets were revealed.

Keywords: *public key, cryptanalysis, post-quantum cryptography, polynomial equations.*

ЗАВАЛИШИНА Елена Владимировна — м.н.с. Института математики им. С.Л.Соболева СО РАН, студентка Новосибирского государственного университета, Новосибирск. E-mail: **e.zavalishina@g.nsu.ru**

УДК 004.75

МЕТОД СОКРЫТИЯ ПРИВАТНЫХ ДАННЫХ ДЛЯ БЛОКЧЕЙН-СИСТЕМЫ ПРОВЕДЕНИЯ ТЕНДЕРОВ[1]

Д. О. Кондырев

Институт математики им. С.Л.Соболева,
Новосибирский государственный университет,
Лаборатория криптографии JetBrains Research, г. Новосибирск, Россия

E-mail: dkondyrev@gmail.com

В работе предложен новый метод, позволяющий решить проблему приватности информации в открытых блокчейн-системах с использованием криптографического протокола доказательства с нулевым разглашением zk-SNARK. Предложенный метод реализован в виде криптографической схемы на основе библиотеки libsnark и интегрирован в модифицированный Ethereum C++ клиент.

Ключевые слова: *тендеры, распределенные системы, блокчейн, доказательство с нулевым разглашением, zk-SNARK, платформа Ethereum.*

На сегодняшний день большинство конкурсных закупок и электронных торгов проводится через специализированные информационные системы. В таких системах участники должны быть уверены в том, что никто не имеет возможности нарушить правила проведения тендера или получить доступ к конфиденциальной информации.

Решить проблему доверия при проведении тендеров позволяет блокчейн. Однако, при использовании этой технологии все данные сохраняются в открытом виде и доступны всем участникам. В случае с тендерами открытость информации нарушает тайну заявок, которая должна быть сохранена до окончания этапа запроса предложений.

Ранее была разработана блокчейн-система для проведения тендеров с шифрованием заявок [1]. Однако такой подход не позволяет проверить корректность зашифрованной заявки в момент ее подачи. Еще одним недостатком является то, что все участники могут наблюдать факт подачи заявки пользователем.

В данной работе предложена и реализована система тендеров, которая удовлетворяет критериям безопасности, открытости и конфиденциальности. Вопрос доверия решен с помощью технологии блокчейн, а сокрытие приватной информации – с помощью алгоритмов доказательства с нулевым разглашением.

Разработанная система основана на платформе Ethereum. Вся ключевая информация о тендерах сохраняется в блокчейне, а проверка правил и отслеживание выполнения условий участниками реализованы в виде кода смарт-контрактов.

В работе предложен новый метод сокрытия приватной информации в открытых блокчейн-системах. Разработанный метод основан на криптографическом протоколе неинтерактивного доказательства знания с нулевым разглашением zk-SNARK [2] и позволяет скрывать конфиденциальную информацию на этапе подачи заявок.

[1]Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации номер 075-15-2019-1613 и лаборатории криптографии JetBrains Research.

Для реализации алгоритма сокрытия информации о заявках в Ethereum C++ клиент был добавлен отдельный модуль *tenderzkp*. Он построен на базе протокола zk-SNARK с предобработкой (preprocessing zk-SNARK) для NP-полного языка системы ограничений ранга 1 (R1CS – rank-1 constraint systems). Этот протокол использует эллиптическую кривую Баррето-Наерига. Реализация этой криптографической схемы предоставлена библиотекой libsnark [3].

В модуле *tenderzkp* реализованы функции для создания и верификации доказательства о корректности заявки. Доказательство строится на основе ограничений на приватные и открытые входные данные заявки, выраженных с помощью базовых схем библиотеки libsnark.

Для работы с добавленной криптографической схемой в Ethereum C++ клиент были созданы новые предкомпилированные контракты с адресами $0x00...09$ и $0x00...0a$. Была разработана Solidity-библиотека, которая инкапсулирует низкоуровневое взаимодействие с предкомпилированными контрактами и предоставляет интерфейс для работы с ними в виде Solidity-функций. Чтобы добавить возможность вызывать методы разработанной криптографической схемы из сторонних приложений был расширен JSON-RPC API Ethereum клиента.

Предложенный и реализованный в данной работе метод может быть использован не только для тендеров, но и в других системах, где есть необходимость скрывать часть информации в открытой блокчейн-сети. Он расширяет область применения технологии блокчейн в промышленных программных комплексах.

ЛИТЕРАТУРА

1. *Hardwick F. S., Akram R. N., and Markantonakis K.* Fair and transparent blockchain based tendering framework — A step towards open governance // IEEE Intern. Conf. TrustCom/BigDataSE, New York, USA, 2018. P. 1342–1347.
2. *Ben-Sasson E., Chiesa A., Genkin D., et al.* SNARKs for C: Verifying program executions succinctly and in zero knowledge // CRYPTO'2013. LNCS. 2013. V. 8043. P. 90–108.
3. `https://github.com/scipr-lab/libsnark` — libsnark: a C++ library for zkSNARK proofs.

Kondyrev D. O. **METHOD OF HIDING PRIVATE DATA FOR THE BLOCKCHAIN TENDER SYSTEM.** A new method has been proposed to solve the problem of information privacy in open blockchain systems using the zk-SNARK cryptographic zero-knowledge proof protocol. The proposed method has been implemented as a cryptographic scheme based on the libsnark library. To integrate the cryptographic scheme into the system, the Ethereum C ++ client has been modified, where new functions and an interface for working with them in the form of precompiled contracts has been added.

Keywords: *tenders, distributed systems, blockchain, zero-knowledge proof, zk-SNARK, Ethereum platform.*

КОНДЫРЕВ Дмитрий Олегович — аспирант факультета информационных технологий Новосибирского государственного университета, исследователь лаборатории криторафии JetBrains Research, младший научный сотрудник Института математики им. С.Л.Соболева, г. Новосибирск. E-mail: **dkondyrev@gmail.com**

УДК 519.7

О МЕТРИЧЕСКИХ СВОЙСТВАХ МНОЖЕСТВА САМОДУАЛЬНЫХ БЕНТ-ФУНКЦИЙ[1]

А. В. Куценко

Бент-функция называется самодуальной, если она совпадает со своей дуальной бент-функцией и анти-самодуальной, — если она совпадает с отрицанием своей дуальной. В данной работе приводится обзор известных метрических свойств множества самодуальных бент-функций. Приводится полный спектр расстояний Хэмминга между самодуальными бент-функциями из класса Мэйорана–МакФарланда. Даются результаты, касающиеся характеризации булевых функций, находящихся на максимально возможном удалении от множества самодуальных бент-функций. Приводится описание групп автоморфизмов множеств самодуальных и анти-самодуальных бент-функций от n переменных. Даётся описание автоморфизмов множества булевых функций от n переменных, которые меняют местами множества самодуальных и анти-самодуальных бент-функций. Приводится описание изометричных отображений, сохраняющих неизменным отношение Рэлея каждой булевой функции от n переменных. Даётся характеризация всех изометричных отображений, сохраняющих максимальную нелинейность и расстояние Хэмминга между каждой бент-функций и дуальной к ней.

Ключевые слова: *булева функция, самодуальная бент-функция, расстояние Хэмминга, изометричное отображение, метрическая регулярность, группа автоморфизмов, отношение Рэлея*

Через $\mathbb{F}_2^n$ обозначим линейное пространство всех двоичных векторов длины n над полем $\mathbb{F}_2$. *Булевой функцией* от n переменных называется отображение вида $\mathbb{F}_2^n \to \mathbb{F}_2$. Множество всех булевых функций от n переменных обозначается через $\mathcal{F}_n$. Для каждой пары $x, y \in \mathbb{F}_2^n$ через $\langle x, y\rangle$ обозначим число $\bigoplus_{i=1}^{n} x_i y_i$, где операция $\oplus$ есть *сложение по модулю* 2. *Весом Хэмминга* $\mathrm{wt}(x)$ вектора $x \in \mathbb{F}_2^n$ называется число его ненулевых координат. *Расстояние Хэмминга* между булевыми функциями f, g от n переменных — число двоичных векторов длины n, на которых эти функции принимают различные значения, обозначается как $\mathrm{dist}(f, g)$. *Преобразованием Уолша — Адамара* булевой функции f от n переменных называется целочисленная функция $W_f : \mathbb{F}_2^n \to \mathbb{Z}$, заданная равенством

$$W_f(y) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus \langle x, y\rangle}, \quad y \in \mathbb{F}_2^n.$$

Булева функция f от чётного числа переменных n называется *бент-функцией*, если $|W_f(y)| = 2^{n/2}$ для каждого $y \in \mathbb{F}_2^n$ [1]. Для множества бент-функций от n переменных используется обозначение $\mathcal{B}_n$. Для каждой $f \in \mathcal{B}_n$ однозначным образом из соотношения $W_f(y) = (-1)^{\widetilde{f}(y)} 2^{n/2}$ определяется *дуальная* к ней бент-функция $\widetilde{f} \in \mathcal{B}_n$, значения которой находятся из соответствия для каждого $y \in \mathbb{F}_2^n$. Бент-функция f называется *самодуальной (анти-самодуальной)*, если $f = \tilde{f}$ (соответственно $f = \widetilde{f} \oplus 1$). Множества самодуальных и анти-самодуальных бент-функций от n переменных обозначаются через $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$, соответственно [2].

[1]Работа выполнена в рамках государственного задания ИМ СО РАН (проект № 0314-2019-0017) при поддержке Российского Фонда Фундаментальных Исследований (проекты № 18-07-01394, 20-31-70043) и лаборатории криптографии JetBrains Research.

Открытой проблемой является полная характеризация и описание класса самодуальных бент-функций. Этому и другим вопросам, связанным с самодуальными бент-функциями, посвящён ряд работ (C. Carlet, L. E. Danielson, M. G. Parker, P. Solé, X. Hou, T. Feulner, L. Sok, A. Wassermann и др.). В частности, в работе [3] приведена аффинная классификация самодуальныех бент-функции от $2, 4, 6$ переменных и всех квадратичных самодуальных бент-функций от 8 переменных относительно преобразования, сохраняющего самодуальность. В статье [2] приведена классификация всех квадратичных самодуальных бент-функций. Аффинную классификацию квадратичных и кубических самодуальных бент-функций от 8 переменных относительно преобразования, сохраняющего самодуальность, можно найти в работе [4]. Верхнюю оценку на количество самодуальных бент-функций можно найти в статье [5]. В работах [6, 7, 8] представлены конструкции самодуальных бент-функций. Связь самодуальных кватернарных бент-функций и самодуальных булевых бент-функций отмечена в [9].

Согласно [10], назовём ортогональной группой порядка n над полем $\mathbb{F}_2$ группу

$$\mathcal{O}_n = \left\{ L \in \mathrm{GL}\,(n, \mathbb{F}_2) : LL^{\mathrm{T}} = I_n \right\},$$

где L^{T} — транспонирование L, и I_n — единичная матрица порядка n над полем $\mathbb{F}_2$.

Далее будут представлены известные результаты, касающиеся метрических свойств самодуальных бент-функций, опубликованные в работах [11, 12, 13] (см. также [14],[15],[16]).

1. Самодуальные бент-функции Мэйорана–МакФарланда

Бент-функции от $2k$ переменных, представимые в виде

$$f(x, y) = \langle x, \pi(y) \rangle \oplus g(y), \quad x, y \in \mathbb{F}_2^k,$$

где π — перестановка на множестве $\mathbb{F}_2^k$, и g — булева функция от k переменных, формируют хорошо известный класс *Мэйорана — МакФарланда* [17]. Данный класс имеет мощность, равную $2^k! \cdot 2^{2^k}$.

Через $\mathrm{SB}^+_{\mathcal{M}}(n)$ обозначим множество самодуальных бент-функций от n переменных из класса Мэйорана — МакФарланда, а через $\mathrm{SB}^-_{\mathcal{M}}(n)$ — множество анти-самодуальных бент-функций от n переменных из класса Мэйорана — МакФарланда. В работе [3] были найдены необходимые и достаточные условия самодуальности бент-функций из класса Мэйорана — МакФарланда, а именно, было показано, что бент-функция $f(x, y)$ Мэйорана — МакФарланда принадлежит множеству $\mathrm{SB}^+_{\mathcal{M}}(2k)$ тогда и только тогда, когда

$$\pi(y) = L(y \oplus c), \quad g(y) = \langle c, y \rangle \oplus d, \quad y \in \mathbb{F}_2^k,$$

где $L \in \mathcal{O}_k$, $c \in \mathbb{F}_2^k$, $\mathrm{wt}(c)$ — чётное число, $d \in \mathbb{F}_2^n$. Заметим, что $\left|\mathrm{SB}^+_{\mathcal{M}}(2k)\right| = 2^k \cdot |\mathcal{O}_k|$.

Всюду далее предполагается, что n — чётное натуральное число. В работе [11] исследовались возможные расстояния Хэмминга между самодуальными бент-функциями из класса Мэйорана — МакФарланда.

Теорема 1 [11]. Пусть $n \geqslant 4$ и $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n) \cup \mathrm{SB}^-_{\mathcal{M}}(n)$, тогда

$$\mathrm{dist}(f, g) \in \left\{ 2^{n-1}, 2^{n-1}\left(1 \pm \frac{1}{2^r}\right), r = 0, 1, ..., n/2 - 1 \right\}.$$

Более того, если $f, g \in \mathrm{SB}^+_{\mathcal{M}}(n)$ или $f, g \in \mathrm{SB}^-_{\mathcal{M}}(n)$, то все приведённые расстояния, кроме 2^{n-1}, являются достижимыми. При этом для произвольной пары функций $f \in \mathrm{SB}^+_{\mathcal{M}}(n)$ и $g \in \mathrm{SB}^-_{\mathcal{M}}(n)$ справедливо $\mathrm{dist}(f, g) = 2^{n-1}$.

Анализ приведённых расстояний позволяет вычислить минимальное расстояние Хэмминга между рассматриваемыми функциями.

Следствие 1. Пусть $n \geqslant 4$, тогда минимальное расстояние Хэмминга между самодуальными бент-функциями от n переменных из класса Мэйорана–МакФарланда равно 2^{n-2}.

Более того, в силу того, что минимальное расстояние Хэмминга между квадратичными булевыми функциями от n переменных (кодовыми словами кода Рида–Маллера $\mathrm{RM}(2,n)$) не меньше, чем 2^{n-2} [18], получаем следующее

Следствие 2. Пусть $n \geqslant 4$, тогда минимальное расстояние Хэмминга между квадратичными булевыми функциями достижимо на самодуальных бент-функциях от n переменных из класса Мэйорана–МакФарланда.

2. Метрическая регулярность

Всюду в этой главе предполагается, что n — чётное натуральное число.

Известно [19], что минимальное расстояние Хэмминга между бент-функциями от n переменных равно $2^{n/2}$. В работе [12] доказано, что при $n \geqslant 4$ данное расстояние достижимо на множестве (анти-)самодуальных бент-функций.

Утверждение 1 [12]. Пусть $n \geqslant 4$, тогда минимальное расстояние Хэмминга между (анти-)самодуальными бент-функциями от n переменных равно $2^{n/2}$.

Пусть $A \subseteq \mathbb{F}_2^n$ — произвольное множество, и $y \in \mathbb{F}_2^n$ — произвольный двоичный вектор. Расстояние от вектора y до множества A определяется как $\mathrm{dist}(y,A) = \min\limits_{x\in A} \mathrm{dist}(y,x)$. *Радиусом покрытия* множества A называется число $\mathrm{d}(A) = \max\limits_{y\in\mathbb{F}_2^n} \mathrm{dist}(y,A)$. Множество двоичных векторов, находящихся на расстоянии $\mathrm{d}(A)$ от множества $A \subseteq \mathbb{F}_2^n$, называется *метрическим дополнением* множества A и обозначается $\widehat{A}$ [20]. Если справедливо $\widehat{\widehat{A}} = A$, то множество называется *метрически регулярным*.

Рассматривая данные определения применительно к векторам значений булевых функций, можно определить *радиус покрытия*, *метрическое дополнение* и *метрическую регулярность* произвольного подмножества $M \subseteq \mathcal{F}_n$ [21].

В работе [3] было доказано, что радиус покрытия множества $\mathrm{SB}^+(n)$ равен 2^{n-1}. Следующее утверждение описывает метрическое дополнение множества самодуальных бент-функций.

Теорема 2 [12]. Пусть $n \geqslant 4$, тогда булева функция от n переменных

— является самодуальной бент-функцией в том и только в том случае, когда она находится на расстоянии 2^{n-1} от множества всех анти-самодуальных бент-функций от n переменных, то есть является элементом множества $\widehat{\mathrm{SB}^-(n)}$;
— является анти-самодуальной бент-функцией в том и только в том случае, когда она находится на расстоянии 2^{n-1} от множества всех самодуальных бент-функций от n переменных, то есть является элементом множества $\widehat{\mathrm{SB}^+(n)}$.

В работе [22] было доказано, что аффинными являются булевы функции, которая находятся на максимально аозможном удалении от множества бент-функций, что влечёт *дуальность* в определении аффинных функций и бент-функций. Таким образом, на основании Теоремы 2 можно говорить о том, что между множествами самодуальных и анти-самодуальных бент-функций от $n \geqslant 4$ переменных существует метрическая *дуальность*.

На основании Теоремы 2 (случай $n = 2$ рассмотрен отдельно) в [12] было показано, что

Следствие 3.

1) Множество $\mathrm{SB}^+(n)$ всех самодуальных бент-функций от n переменных является метрически регулярным;
2) Множество $\mathrm{SB}^-(n)$ всех анти-самодуальных бент-функций от n переменных является метрически регулярным.

3. Группа автоморфизмов

Отображение всех булевых функций от n переменных в себя называется *изометричным*, если оно сохраняет расстояние Хэмминга между каждой парой булевых функций от n переменных. Множество изометричных отображений множества всех булевых функций от n переменных в себя будем обозначать через $\mathcal{I}_n$. Известно, что каждое такое отображение однозначно представляется в виде

$$f(x) \longrightarrow f(\pi(x)) \oplus g(x),$$

где π — перестановка на множестве $\mathbb{F}_2^n$, а g — булева функция от n переменных [23]. Отображение такого вида обозначм через $\varphi_{\pi,g} \in \mathcal{I}_n$. Известно, что каждое изометричное отображение множества всех булевых функций от чётного числа переменных n в себя, оставляющее множество $\mathcal{B}_n$ на месте, представимо в виде композиция аффинного преобразования координат и прибавления аффинной функции от n переменных [24].

Группой автоморфизмов фиксированного подмножества $M \subseteq \mathcal{F}_n$ называется группа элементов множества $\mathcal{I}_n$, оставляющая множество M на месте. Группа автоморфизмом множества M обозначается через $\mathrm{Aut}\,(M)$.

Далее предполагается, что n — чётное натуральное число.

В работе [4] (см. также [3]) было доказано, что отображение всех булевых функций от n переменных в себя, имеющее вид

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x\rangle \oplus d,$$

где $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ — чётное число, $d \in \mathbb{F}_2$, сохраняет самодуальность бент-функции. Нетрудно видеть, что все отображения данного вида являются элементами множества $\mathcal{I}_n$. Группа таких преобразований называется *расширенной ортогональной группой* и обозначается $\overline{\mathcal{O}}_n$ [4, 25]. Известно, что $\overline{\mathcal{O}}_n$ является подгруппой группы $\mathrm{GL}\,(n+2, \mathbb{F}_2)$ [4].

В работе [2] было отмечено, что отображение всех булевых функций от n переменных в себя, имеющее вид

$$f(x) \longrightarrow f\left(x \oplus c\right) \oplus \langle c, x\rangle\,,$$

где $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ — нечётное число, определяет биекцию между множествами $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$. Очевидно, что такое отображение сохраняет расстояние Хэмминга. Частный случай отображения данного вида — при $c = (1, 0, 0, ..., 0) \in \mathbb{F}_2^n$ — ранее был расмотрен в статье [3], на основании чего был сделан вывод о том, что между множествами $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$ существует взаимно-однозначное соответствие.

В статье [13] получено обобщение данных результатов в рамках класса изометричных отображений. Было доказано, что группы автоморфизмов множеств $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$ совпадают.

Теорема 3 [13]. При $n \geqslant 4$, справедливо $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right)$.

Был получен следующий критерий сохранения самодуальности.

Теорема 4 [13]. Пусть $n \geqslant 4$, тогда изометричное отображение $\varphi_{\pi,g}$ является элементом гриппы $\mathrm{Aut}\left(\mathrm{SB}^+(n)\right)$ в том и только в том случае, когда для любых $x, y \in \mathbb{F}_2^n$ справедливо

$$\langle \pi(x), y\rangle \oplus g(x) = \left\langle x, \pi^{-1}(y)\right\rangle \oplus g\left(\pi^{-1}(y)\right).$$

С использованием вышеупомянутого критерия и Теоремы 3 было получено описание группы автоморфизмов множества (анти-)самодуальных бент-функций от n переменных.

Теорема 5 [13]. При $n \geqslant 4$ справедливо

$$\mathrm{Aut}\left(\mathrm{SB}^+(n)\right) = \mathrm{Aut}\left(\mathrm{SB}^-(n)\right) = \overline{\mathcal{O}}_n.$$

Из полученных результатов следует, что более общего подхода к классификации самодуальных бент-функций на основе изометричных отображений, чем предложенный в работах [3, 4], не существует.

Применительно к биекциям между множествами $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$ был получен следующий критерий.

Теорема 6 [13]. Пусть $n \geqslant 4$, тогда изометричное отображение $\varphi_{\pi,g}$ определяет биекцию между множествами $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$ в том и только в том случае, когда для любых $x, y \in \mathbb{F}_2^n$ справедливо

$$\langle \pi(x), y\rangle \oplus g(x) = \left\langle x, \pi^{-1}(y)\right\rangle \oplus g\left(\pi^{-1}(y)\right) \oplus 1.$$

С использованием данного критерия была получена общая форма изметричных отображений, определяющих биекцию между множествами $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$.

Теорема 7 [13]. При $n \geqslant 4$ изометричное отображение $\varphi_{\pi,g} \in \mathcal{I}_n$ определяет биекцию между множествами $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$ если и только если

$$\pi(x) = L\left(x \oplus c\right),\ g(x) = \langle c, x\rangle \oplus d,\ x \in \mathbb{F}_2^n,$$

где $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ — чётное число, $d \in \mathbb{F}_2$.

Из Теорем 5 и 7 следует, что чётность веса Хэмминга вектора $c \in \mathbb{F}_2^n$, фигурирующего в описании расширенной ортогональной группы, является "переключателем" между изометричным отображением, сохраняющим (анти-)самодуальность, и изометричным отображением, меняющим местами самодуальные и анти-самодуальные бент-функции.

4. Расстояние Хэмминга между бент-функций и дуальной к ней

Согласно [3, 25] *отношением Рэлея* (the Rayleigh quotient) S_f булевой функции f от n переменных называется число

$$S_f = \sum_{x,y\in\mathbb{F}_2^n} (-1)^{f(x)\oplus f(y)\oplus\langle x,y\rangle} = \sum_{y\in\mathbb{F}_2^n} (-1)^{f(y)} W_f(y).$$

Известно [3], что абсолютное значение S_f не превосходит числа $2^{3n/2}$, при этом в случае, когда n — чётное число, данная оценка достигается только на самодуальных бент-функциях $\left(+2^{3n/2}\right)$ и анти-самодуальных бент-функциях $\left(-2^{3n/2}\right)$.

Далее предполагается, что n — чётное натуральное число.

В работе [13] были исследованы вопросы сохранения, а также смены знака отношения Рэлея каждой булевой функции от n переменых при изометричных преобразованиях.

Теорема 8 [13]. Пусть $n \geqslant 4$, тогда изометричное отображение $\varphi_{\pi,g} \in \mathcal{I}_n$ сохраняет отношение Рэлея каждой булевой функции от n переменных в том и только в том случае, когда $\varphi_{\pi,g} \in \mathrm{Aut}\left(\mathrm{SB}^+(n)\right)$.

Теорема 9 [13]. Пусть $n \geqslant 4$, тогда изометричное отображение $\varphi_{\pi,g} \in \mathcal{I}_n$ меняет знак отношения Рэлея каждой булевой функции от n переменных в том и только в том случае, когда оно определяет биекцию между множествами $\mathrm{SB}^+(n)$ и $\mathrm{SB}^-(n)$.

Пусть $f \in \mathcal{B}_n$, из соотношения

$$\mathrm{dist}(f, \widetilde{f}) = 2^{n-1} - \frac{1}{2^{n/2+1}} S_f$$

следует, что отношение Рэлея полностью характеризует расстояние Хэмминга между бент-функцией $f \in \mathcal{B}_n$ и дуальной к ней функцией $\widetilde{f} \in \mathcal{B}_n$. Таким образом, на основе Теорем 5 и 8 можно получить следующий результат.

Теорема 10 [13]. При $n \geqslant 4$ изометричное отображение $\varphi_{\pi,g} \in \mathcal{I}_n$ оставляет множество бент-функций от n переменных на месте и сохраняет расстояние Хэмминга между бент-функцией и дуальной к ней тогда и только тогда, когда

$$f(x) \longrightarrow f\left(L\left(x \oplus c\right)\right) \oplus \langle c, x \rangle \oplus d, \quad x \in \mathbb{F}_2^n, \tag{1}$$

где $L \in \mathcal{O}_n$, $c \in \mathbb{F}_2^n$, $\mathrm{wt}(c)$ — чётное число, $d \in \mathbb{F}_2$.

ЛИТЕРАТУРА

1. *Rothaus O.* On bent functions // J. Combin. Theory. Ser. A, 1976, vol. 20, no. 3, pp. 300–305.
2. *Hou X.-D.* Classification of self dual quadratic bent functions // Des. Codes Cryptogr., 2012, vol. 63, no. 2, pp. 183–198.
3. *Carlet C., Danielson L. E., Parker M. G., Solé P.* Self-dual bent functions. // Int. J. Inform. Coding Theory, 2010, vol. 1, pp. 384–399.
4. *Feulner T., Sok L., Solé P., Wassermann A.* Towards the classification of self-dual bent functions in eight variables // Des. Codes Cryptogr., 2013, vol. 68, no. 1, pp. 395–406.
5. *Hyun J. Y., Lee H., Lee Y.* MacWilliams duality and Gleason-type theorem on self-dual bent functions // Des. Codes Cryptogr., 2012, vol. 63, no. 3, pp. 295–304.
6. *Luo G., Cao X., Mesnager S.* Several new classes of self-dual bent functions derived from involutions // Cryptogr. Commun., 2019, vol. 11, no. 6, pp. 1261–1273.
7. *Mesnager S.* Several new infinite families of bent functions and their duals // IEEE Trans. Inf. Theory, 2014, vol. 60, no. 7, pp. 4397–4407.
8. *Rifà J., Zinoviev V. A.* On binary quadratic symmetric bent and almost bent functions // arXiv:1211.5257v3, 2019.
9. *Sok L., Shi M., Solé. P.* Classification and Construction of quaternary self-dual bent functions // Cryptogr. Commun., 2018, vol. 10, no. 2, pp. 277–289.
10. *Janusz G. J.* Parametrization of self-dual codes by orthogonal matrices // Finite Fields Appl., 2007, vol. 13, no. 3, pp. 450—491.
11. *Kutsenko A. V.* The Hamming distance spectrum between self-dual Maiorana-McFarland bent functions // J. Appl. Industr. Math., 2018, vol. 12, no. 1, pp. 112–125.
12. *Kutsenko A.* Metrical properties of self-dual bent functions // Des. Codes Cryptogr., 2020, vol. 88, no. 1, pp. 201–222.
13. *Kutsenko A.* The group of automorphisms of the set of self-dual bent functions // Cryptogr. Commun., 2020. DOI: 10.1007/s12095-020-00438-y

14. *Куценко А. В.* О множестве расстояний Хэмминга между самодуальными бент-функциями // Прикладная дискретная математика. Приложение. 2016, № 9, С. 29–30. DOI: 10.17223/2226308X/9/11
15. *Куценко А. В.* О некоторых свойствах самодуальных бент-функций // Прикладная дискретная математика. Приложение. 2018, № 11, С. 44–46. DOI: 10.17223/2226308X/11/13
16. *Куценко А. В.* Изометричные отображения множества всех булевых функций в себя, сохраняющие самодуальность и отношение Рэлея // Прикладная дискретная математика. Приложение. 2019, № 12, С. 55–58. DOI: 10.17223/2226308X/12/16
17. *McFarland R. L.* A family of difference sets in non-cyclic groups // J. Combin. Theory. Ser. A, 1973, vol. 15, no. 1, pp. 1–10.
18. *MacWilliams F. J., Sloane N. J. A.* The Theory of Error-Correcting Codes. Amsterdam, New York, Oxford, North-Holland, 1983. 782 p.
19. *Коломеец Н. А., Павлов А. В.* Свойства бент-функций, находящихся на минимальном расстоянии друг от друга // Прикладная дискретная математика. 2009. № 4 С. 5–20.
20. *Облаухов А. К.* О метрическом дополнении подпространств булева куба // Дискретн. анализ и исслед. опер., 2016, вып. 23, № 3. (2016), С. 93–106
21. *Tokareva N.* Bent Functions: Results and Applications to Cryptography. Acad. Press, Elsevier, 2015. 230 p.
22. *Tokareva N.* Duality between bent functions and affine functions. Discrete Math., 2012, vol. 312, no. 3, pp. 666–670.
23. *Марков А. А.* О преобразованиях, не распространяющих искажения, Избранные труды. Т. II. Теория алгорифмов и конструктивная математика, математическая логика, информатика и смежные вопросы, МЦНМО, Москва, (2003), 70–93.
24. *Tokareva N. N.* The group of automorphisms of the set of bent functions // Discrete Mathematics and Applications, 2010, vol. 20, no. 5, pp. 655–664.
25. *Danielsen L. E., Parker M. G., Solé P.* The Rayleigh quotient of bent functions // LNCS, 2009, vol. 5921, pp. 418–432.

Kutsenko A. V. **ON METRICAL PROPERTIES OF THE SET OF SELF-DUAL BENT FUNCTIONS.** For every bent function f its dual bent function $\tilde{f}$ is uniquely defined. If $\tilde{f} = f$ then f is called *self-dual bent* and it is called *anti-self-dual bent* if $\tilde{f} = f \oplus 1$. In this work we give a review of metrical properties of the set of self-dual bent functions. We give a complete Hamming distance spectrum between self-dual Maiorana — McFarland bent functions. The set of Boolean functions which are maximally distant from the set of self-dual bent functions is discussed. We give a characterization of automorphim groups of the sets of self-dual and anti-self-dual bent functions in n variables as well as the description of isometric mappings that define bijections between the sets of self-dual and anti-self dual bent functions. The set of isometric mappings which preserve the Rayleigh quotient of a Boolean function is given. As a corollary all isometric mappings which preserve bentness and the Hamming distance between bent function and its dual are given.
Keywords: *Boolean function, self-dual bent function, Hamming distance, isometric mapping, metrical regularity, automorphism group, Rayleigh quotient of Sylvester Hadamard matrix*

КУЦЕНКО Александр Владимирович — аспирант механико-математического факультета Новосибирского национального исследовательского государственного университета, Институт математики им. С. Л. Соболева СО РАН, г. Новосибирск. E-mail: **AlexandrKutsenko@bk.ru**

УДК 519.7

КРИПТОГРАФИЧЕСКИЕ СВОЙСТВА ОРТОМОРФИЗМОВ[1]

Ю. П. Максимлюк

Институт математики им. С. Л. Соболева, г. Новосибирск, Россия

E-mail: yumaximlyuk@gmail.com

В работе рассмотрены взаимно однозначные отображения $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$, называемые ортоморфизмами, такие, что отображения $G(x) = F(x) \oplus x$ так же являются взаимно однозначными. Они используются в схеме Лая-Месси в качестве перемешивающего элемента между раундами, а также для построения криптографически стойких S-блоков. Исследовались основные криптографические свойства, а именно, нелинейные характеристики и дифференциальная равномерность. Выявлено, что ортоморфизмы от малого числа переменных не устойчивы к линейному и дифференциальному криптоанализам.

Ключевые слова: *ортоморфизм, таблица линейного преобладания, таблица дифференциалов.*

В симметричной криптографии часто используются отображения множества $\mathbb{Z}_2^n$, состоящего из двоичных наборов длины n, на себя. В частности, в книге [1] в шифрах FOX (IDEA NXT), использующих схему Лая-Месси, предлагается использовать отображение, называемое ортоморфизмом.

Ортоморфизм $\mathbb{Z}_2^n$ – это взаимно однозначное отображение $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ такое, что отображение $G(x) = F(x) \oplus x$ так же является взаимно однозначным, где $\oplus$ – побитовое сложение по модулю 2.

В литературе в основном освещаются перемешивающие свойства ортоморфизмов. Например, в работе [2] ортоморфизмы характеризуются свойством отображать каждую максимальную подгруппу группы двоичных наборов длины n наполовину в себя и наполовину в свое дополнение.

В рамках данной работы была написана программа, которая использует разработанный рекурсивный алгоритм построения всех ортоморфизмов для заданного n. Она перебирает все значения для k-го элемента и проверяет выполнение определения ортоморфизма. Если проверка успешна, то переходит к $k+1$-му элементу, иначе проверяется следующее значение k-го элемента. Когда проверены все возможные значения для k-ой позиции, происходит возврат к дальнейшей проверки значений для позиции $k-1$. С помощью этой программы были получены ортоморфизмы для малых значений n и один отроморфизм при $n = 16$ для исследования модификации шифра Simon 32/64 [3], где вместо сети Фейстеля использовалась схема Лая-Месси.

Для малых значений n было получено, что:

- при $n = 2$ существует 8 ортоморфизмов;
- при $n = 3$ существует 384 ортоморфизма;
- при $n = 4$ существует 244744192 ортоморфизма.

[1]Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации номер 075-15-2019-1613 и лаборатории криптографии JetBrains Research.

Для всех полученных ортоморфизмов эксперементально исследовались основные криптографические свойства, а именно, нелинейные характеристики и дифференциальная равномерность.

Обозначим вход и выход функции $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ через $x = (x_1, \ldots, x_n)$ и $y = (y_1, \ldots, y_n)$ соответственно. Для линейного криптоанализа строится таблица линейного преобладания, где на пересечении строки $u \in \mathbb{Z}_2^n$ и столбца $v \in \mathbb{Z}_2^n$ находится число λ такое, что соотношение $\langle u, x\rangle = \langle v, y\rangle$ выполняется с вероятностью $(2^{n-1} + \lambda)/2^n$, где $\langle u, x\rangle = u_1x_1 \oplus \cdots \oplus u_nx_n$.

Утверждение 1. При n равных 2, 3 и 4 таблицы линейного преобладания ортоморфизмов состоят из значений 0 и $\pm 2^{n-1}$.

Для дифференциального криптоанализа в таблице дифференциалов на пересечении строки $u \in \mathbb{Z}_2^n$ и столбца $v \in \mathbb{Z}_2^n$ находится число λ такое, что равенство $F(x \oplus u) \oplus F(x) = v$ выполняется в точности для λ различных x.

Утверждение 2. При n равных 2, 3 и 4 таблицы дифференциалов ортоморфизмов состоят из значений 0 и 2^n.

Для полученного ортоморфизма при $n = 16$ также исследовались таблицы линейного преобладания и дифференциалов. Таблица линейного преобладания состоит из значений 0 и $\pm 2^{n-1}$, а таблица дифференциалов из 0 и 2^n.

Утверждения 1, 2 и точечное исследование ортоморфизма для $n = 16$ позволяют предположить, что для любого значения n таблицы дифференциалов и линейного преобладания ортоморфизмов имеют вид, описанный выше. Из чего следует, что ортоморфизмы сами по себе не устойчивы к линейному и дифференциальному криптоанализам и должны использоваться в шифрах в качестве вспомогательного элемента – для построения более устойчивых к криптоанализу перемешивающих отображений.

ЛИТЕРАТУРА

1. *Nakahara Jr. J.* Lai-Massey Cipher Designs. History, Design Criteria and Cryptanalysis // Springer Nature Switzerland AG, 2018.
2. *Mittenthal L.* Block Substitutions Using Orthomorphic Mappings // Advances in Applied Mathematics 16, 59-71, 1995.
3. *Beaulieu R., Shors D., Smith J., Treatman-Clark S., Weeks B., Wingers L.* The Simon and Speck Families Of Lightweight Block Ciphers // Cryptology ePrint Archive, Report 2013/404, 2013

Maksimluk J. P. **CRYPTOGRAPHIC PROPERTIES OF ORTHOMORPHIC PERMUTATIONS.** In this paper, we consider bijective mappings $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ called orthomorphisms such that the mappings $G(x) = F(x) \oplus x$ are also bijective. It is used in the Lai-Massey scheme as a mixing element between rounds and it also can be used to construct cryptographically strong S-boxes. The main cryptographic properties are studied, namely nonlinearity and differential uniformity. It turned out that orthomorphisms of a small number of variables are not resistant to linear and differential cryptanalysis.
Keywords: *orthomorphic permutation, linear approximation table, difference distribution table.*

МАКСИМЛЮК Юлия Павловна — м.н.с. Института математики им. С. Л. Соболева СО РАН, студентка магистратуры Новосибирского государственного университета, исследователь лаборатории криптографии JetBrains Research, Новосибирск. E-mail: **yumaximlyuk@gmail.com**

УДК 519.7

О РАЗЛОЖЕНИИ ВЕКТОРНОЙ БУЛЕВОЙ ФУНКЦИИ В КОМПОЗИЦИЮ ДВУХ ВЕКТОРНЫХ ФУНКЦИЙ[1]

Г. М. Пинтус

В данной работе исследуется возможность представления векторной булевой функции в виде композиции двух векторных булевых функций меньшей алгебраической степени. Вводится понятие разложимости векторной булевой функции. Изучен вопрос сохранения разложимости при расширенном аффинном преобразовании. Представлена конструкция векторной булевой функции третьей степени от произвольного числа переменных, являющейся разложимой. Также был проведен вычислительный экперимент, в результате которого было доказано, что все кубические векторные булевы функции от трёх переменных являются разложимыми.

Ключевые слова: *векторная булева функция, декомпозиция, пороговая реализация*

Атаки по сторонним каналам [1] — вид атак, целью которых является нахождение уязвимостей в реализации криптографической системы. На данный момент эти атаки являются одними из наиболее эффективных среди всех видов криптоанализа. В атаках по сторонним каналам используется информация, полученная при отслеживании перепадов напряжений, времени выполнения процессов, электромагнитного излучения или звуков при проводимых алгоритмом вычислениях.

Пороговая реализация [2] является контрмерой по отношению к атакам по сторонним каналам, разделяя наборы входных данных и используемые векторные булевы функции на части, позволяя скрыть различия между операциями. Таким образом, если разбиение удовлетворяет ряду условий, при работе алгоритма не происходит утечки информации, которая может быть использована в атаке по сторонним каналам.

В данном методе необходимо построить разбиение для векторной булевой функции определенным образом, что не всегда удается сделать. Однако был придуман способ решения данной проблемы, использующий построение разбиения для более простых функций, композицией которых является изначально рассматриваемая векторная булева функция.

В данной работе анализируется возможность представления векторных булевых функций в виде композиции векторных булевых функций меньших степеней. В первую очередь данная задача применима к пороговой реализации, которую не всегда возможно осуществить с изначальной векторной булевой функцией из алгоритма, но возможно с функциями, композиция которых равняется данной. Были рассмотрены векторные булевы функции от трех переменных с алгебраической степенью равной трем и возможность их представления в виде композии двух векторных булевых функций алгебраической степени два.

Так как важным при рассмотрении является сохранение свойств при преобразованиях, а одним из наиболее распространенных является расширенное аффинное преобразование, мы исследуем вопрос сохранения разложимости векторной булевой функции при расширенной аффинной эквивалентности.

Векторной булевой функцией ((n,m)-функцией) F называется произвольное отображение $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. В случае $m = 1$ говорят, что F —- *булева функция от* n

[1]Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

переменных. Векторная булева (n, m)-функция F может быть задана набором из m координатных булевых функций от n переменных: $F(x) = (f_1(x), f_2(x), ..., f_m(x)), x \in \mathbb{F}_2^n$. Любую (n, m)- функцию можно единственным образом записать в виде *полинома Жегалкина*, или *алгебраической нормальной формы* (АНФ):

$$F(x_1, ..., x_n) = \left(\bigoplus_{k=1}^{n} \bigoplus_{i_1,...,i_k} a_{i_1,...,i_k} x_{i_1} \cdot ... \cdot x_{i_k} \right) \oplus a_0,$$

где для каждого k индексы $i_1, ..., i_k$ попарно различны и множества $\{i_1, ..., i_k\}$ являются всеми различными непустыми подмножествами множества $\{1, ..., n\}$; коэффициенты $a_{i_1,...,i_k}, a_0$ принимают значения из $\mathbb{F}_2^m$. *Алгебраической степенью* $\deg(F)$ функции $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ называется количество переменных в самом длинном слагаемом АНФ, при котором коэффициент не равен нулевому вектору. Функция степени не выше 1 называется *аффинной*, при этом в случае $a_0 = 0$ функция *линейна*.

Две векторные (n, n)-функции F и G называются *расширенно аффинно эквивалентными* (EA-эквивалентными), если существуют две аффинные (n, n)-подстановки A, B на множестве $\mathbb{F}_2^n$ и аффинная (n, n)-функция C, такие что $G(x) = (B \circ F \circ A)(x) + C(x), x \in \mathbb{F}_2^n$.

Пусть F — векторная булева (n, n)-функция, такая, что существуют векторные булевы (n, n)-функции G, H, такие что $\max\{\deg(G), \deg(H)\} < \deg(F)$ и $F(x) = G(H(x))$ для всех $x \in \mathbb{F}_2^n$. Векторную булеву (n, n)-функцию F степени $d > 2$, допускающую такую декомпозицию, будем называть *разложимой*.

Теорема 1. Пусть (n, n)-функция F степени $d > 2$ разложима. Тогда (n, n)-функция $F' = A_2 \circ F \circ A_1$, где A_1, A_2 — произвольные аффинные (n, n)-подстановки, также будет являться разложимой. Если F представима в виде композиции двух (n, n)-функций G, H степениньше d,таких что функция H обратима, и для функции H^{-1} справедливо условие $\deg(H^{-1}) \leqslant \max\{\deg(G), \deg(H)\}$, то (n, n)-функция $F'' = F + A_0$ будет разложимой для любой аффинной (n, n)-функции A_0 .

Также была получена конструкция, которая позволяет для любого n построить класс разложимых векторных булевых функций третьей степени.

Теорема 2. Пусть $i, j, p, q \in \{1, .., n\}$ — числа, такие что $i \neq j$ и $p \neq q$, $\{l_k\}_{k=1}^n, \{l'_r\}_{r=1}^n$ — наборы произвольных линейных булевых функций от n переменных, такие что $\deg(x_p x_q(l_i(x) + l_j(x))) = 3$, $Y(x) = (y_1(x), ..., y_n(x))$, где $y_k(x) = x_p x_q + l_k(x)$ при $k = 1, \ldots, n$, $x \in \mathbb{F}_2^n$. Тогда разложимой является векторная булева функция $F(x)$, определенная следующим образом:

$$F(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ ... \\ f_n(x) \end{pmatrix} = \begin{pmatrix} x_p x_q(l_i(x) + l_j(x)) + x_p x_q + l_i(x) l_j(x) + l'_1(Y(x)) \\ x_p x_q(l_i(x) + l_j(x)) + x_p x_q + l_i(x) l_j(x) + l'_2(Y(x)) \\ ..... \\ x_p x_q(l_i(x) + l_j(x)) + x_p x_q + l_i(x) l_j(x) + l'_n(Y(x)) \end{pmatrix}$$

ЛИТЕРАТУРА

1. *Bhunia S., Tehranipoor M.* Side-Channel Attacks // Hardware Security. A Hands-On Learning Approach. 2019. P 193—218.
2. *Nikova S., Rechberger C., Rijmen V.* Threshold Implementations Against Side-Channel Attacks and Glitches // Information and Communication Technologies. 2006. N. 4307 P 529—546.

Pintus G.M. **ON DECOMPOSITION OF VECTORIAL BOOLEAN FUNCTION IN COMPOSITION WITH TWO VECTORIAL FUNCTIONS .** The condition of

preserving the possibility of representation the vectorial Boolean function as a composition of two vectorial Boolean functions of lower degrees after extended affine transformation was proved. The construction of a vectorial Boolean function of the third degree from an arbitrary number of variables that is decomposable is presented. Also a computational experiment was conducted, which proved that all vector Boolean functions of the third degree in three variables are decomposable.
Keywords: *vectorial boolean function, decomposition, threshold implementation*

ПИНТУС Георгий Михайлович — студент Новосибирского государственного университета, Новосибирск. E-mail: **g.pintus@g.nsu.ru**

УДК 519.7

О ПРИМЕНЕНИИ SAT-РЕШАТЕЛЕЙ В КРИПТОАНАЛИЗЕ[1]

Д. А. Софронова, К. В. Калгин

Новосибирский государственный университет, Лаборатория криптографии JetBrains Research, Новосибирск, Россия

Институт математики имени С.Л.Соболева

E-mail: d.sofronova1@g.nsu.ru, Kalginkv@gmail.com

В работе представлен программный комплекс, позволяющий преобразовывать описание криптографической задачи (шифр, хэш-функция, поиск криптографических функций) в КНФ. В дальнейшем SAT-решатель устанавливает истинность формулы и находит означивание переменных. Отличительные особенности данной разработки - универсальность, малый объем исходного кода (300 строк C++), легко модифицируемая и расширяемая реализация.

Ключевые слова: *криптоанализ, SAT-решатель, атака "угадай-и-вычисли".*

В основе одного из методов анализа симметричных шифров лежит использование SAT-решателей. Для этого исходный шифр или хэш-функция записываются в виде логической формулы, истинность которой предстоит установить SAT-решателю. SAT-задача – задача определения выполнимости логической формулы [4]. SAT-решатель – программа, которая ищет означивание переменных, на котором формула истинна. Логическая формула записывается в конъюнктивной нормальной форме (конъюнкция дизъюнкций переменных, далее – КНФ). Известно, что эта задача NP-полная. Существует ли в общем случае алгоритм поиска подходящего значения переменных не известно. Поиск начальных значений, подающихся исследуемому алгоритму, которые являются решением поставленных задач, производится SAT-решателем. SAT-решатель не гарантирует, что означивание найдется за полиномиальное время. Однако для многих практических задач такой подход позволяет определять выполнимость формул с тысячами переменных. Для проведения криптоанализа с помощью SAT-решателя необходим только механизм для представления криптографических алгоритмов в виде КНФ в формате DIMACS.

На данный момент существуют две разработки с разным подходом, добившиеся хороших результатов в решении проблемы автоматизации криптоанализа: Grain of salt [1] и Transalg [2].

Transalg универсален и позволяет сводить к задаче о выполнимости не только криптографические задачи, но и некоторые задачи биоинформатики. Описание шифров происходит на специальном си-подобном языке с последующей генерацией КНФ . Уже реализованы ряд шифров и хэш-функций [5]. Являясь полноценным транслятором, Transalg анализирует текст описания с помощью лексического, синтаксического и семантического анализаторов, что делает его достаточно сложным для модификации и расширения.

Grain of Salt (далее – GoS) — программный комплекс описания поточных шифров и последующего автоматического проведения атаки "угадай-и-вычисли", который

[1]Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации номер 075-15-2019-1613 и лаборатории криптографии JetBrains Research.

разработал автор cryptominisat [3], M. Soos. Данный вариант хорошо оптимизирован с помощью карт Карно, espresso (логический оптимизатор), предсказыванием значений переменных, поэтому выходная КНФ имеет меньший размер и упрощает работу SAT-решателя. GoS предназначен для описания поточных шифров, построенных на базе регистров сдвига. Другая важная особенность — автоматизация проведения атаки "угадай-и-вычисли" на шифр. Автором [1] уже реализованы шифры Grain, Trivium, Bivium, Crypto1 и Hitag2. Данная разработка позволяет описать только шифры, основанные на регистрах сдвига и фильтрующих функциях, другие не могут быть представлены в этом программном комплексе (например, A5/1 из-за отсутствия поддержки if/else конструкции, другие симметричные шифры и хэш-функции).

В данной работе представлет программный комплекс, одновременно универсальный, легко расширяемый, простой и понятный для пользователей (в том числе на уровне реализации). Под криптографическими задачами далее подразумеваем не только задачи анализа шифров и хэш-функций, но и задачи поиска криптографических функций, определения эквивалентности булевых и векторных функций.

Основная идея заключается в том, что криптографическая задача (алгоритм или множество ограничений) описывается на языке C++ с использованием специальных классов varBool и varInt, у которых переопределены все операторы. Полиморфизм в C++ позволяет переопределить работу операторов для новых типов так, что при выполнении некоторых действий над данными происходит формирование КНФ (в зависимости от операций добавляются разные конструкции) или же реальное исполнение алгоритма. С помощью параметров настраивается результат работы – вычисление выходного значения или заполнение КНФ. Также есть возможность отметить константность определенного значения varBool для оптимизации выходной КНФ. Работа программы построена на операциях, обрабатывающих новые типы и неявно формирующих КНФ на основе логики операций. Это обеспечивает простоту работы с системой, возможность посмотреть код программы и специализировать описание под свои конкретные задачи. Немаловажным плюсом является то, что большинство шифров описываются на языке C. Криптоанализ таких шифров легко осуществляется в проекте заменой типов данных в коде. Также этот факт позволяет утверждать правильность работы ядра, ведь отладить код можно на примерах входных и выходных данных шифров, предоставленных их создателями. Программа является гибкой, использование всех возможностей языка C++ делает реализованный предметно-ориентированный язык крайне функциональным – циклы, условные операторы, шаблоны — все это позволяет описывать алгоритмы разной сложности. Также необходимо отметить возможность означивания переменных или реализации других условий на любом шаге описания задачи. В качестве дополнительных возможностей уже описан механизм регистров линейного сдвига с определенными методами, позволяющих реализацию алгоритмов, основанных на этой технологии. Реализованы шифры, приведенные в описании GoS и шифр A5/1.

ЛИТЕРАТУРА

1. *Soos M. Grain of salt—an automated way to test stream ciphers through SAT solvers // Tools, vol 10, pp. 131-144, 2010*
2. *Otpuschennikov I., Semenov A., Gribanova I., Zaikin O., Kochemazov S. Encoding Cryptographic Functions to SAT Using TRANSALG System// ECAI 2016, Frontiers in Artificial Intelligence and Applications, vol 285, IOS Press, 2016.*

3. *Soos M., Nohl K., Castelluccia C. Extending SAT Solvers to Cryptographic Problems. // Theory and Applications of Satisfiability Testing - SAT 2009. LNCS, vol 5584. Springer, Berlin, Heidelberg*
4. *A. Biere, M. Heule, H. Maaren, T. Walsh, "HANDBOOK OF SATISFIABILITY//Frontiers in Artificial Intelligence and Applications". IOS Press, 2009.*
5. *https://gitlab.com/satencodings/satencodings/*

Sofronova D.A., Kalgin K.V. **ABOUT SAT-SOLVERS IN CRYPTOANALYSIS.** The paper presents a program that allows you to convert the description of a cryptographic task to CNF. A SAT solver establishe the truth of the formula and finds the meaning of the variables after that. Features of this development are universality, a small code size (300 lines of C ++), an easily modifiable and extensible implementation.
Keywords: *cryptanalysis, SAT-solver, "guess-and-determine".*

КАЛГИН Константин Викторович — к.ф.-м.н., м.н.с. Института математики им.С.Л.Соболева, м.н.с. Института вычислительной математики и математической геофизики СО РАН, старший преподаватель Новосибирского государственного университета, исследователь лаборатории криптографии JetBrains Research. E-mail: **kalginkv@gmail.com**

СОФРОНОВА Дарья Алексеевна — студентка Новосибирского государственного университета, Новосибирск, исследователь лаборатории криптографии JetBrains Research. E-mail: **d.sofronova1@g.nsu.ru**

УДК 519.7

ОЦЕНКА НЕЛИНЕЙНОСТИ СБАЛАНСИРОВАННЫХ БУЛЕВЫХ ФУНКЦИЙ, ПОРОЖДЕННЫХ ОБОБЩЕННОЙ КОНСТРУКЦИЕЙ ДОББЕРТИНА[1]

И. А. Сутормин

Новосибрсий государственный университет, г. Новосибирск, Россия;
Институт математики им. С. Л. Соболева, г. Новосибирск, Россия

E-mail: ivan.sutormin@gmail.com

В работе предложено обобщение конструкции Доббертина 1995 г. для высоконелинейных сбалансированных булевых функций. Исследован спектр Уолша-Адамара и получены оценки спкетрального радиуса получившихся функций. Доказана точная верхняя оценка на спектральный радиус (нижняя оценка нелинейности), и предложен способ построить сбалансированную функцию от $2n$ переменных при помощи сбалансированной θ от $n-k$ переменных со спектральным радиусом $2^n + 2^k R_\theta$, где R_θ - спектральный радиус θ.

Ключевые слова: *булевы функции, бент-функции, сбалансированность, нелинейность, спектральный радиус*

В различных криптографических алгоритмах часто используются булевы функции. Нелинейность — одно из основных для них свойств. Оно показывает, насколько хорошо функцию можно приблизить некоторой линейной функцией, работать с которой значительно проще. Шифр может стать уязвимым к линейному криптоанализу при низкой нелинейности даже одной его части. Примером криптографического алгоритма, скомпрометированного своими компонентами с низкой нелинейностью, может послужить старый стандарт шифрования США — DES.

Введём необходимые определения. *Преобразование Уолша–Адамара* булевой функции f определяется как $W_f(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+\langle x,a \rangle}$, $a \in \mathbb{F}_2^n$, *спектральный радиус* $R_f = \max\limits_{a \in \mathbb{F}_2^n} |W_f(a)|$ и *нелиненость* $N_f = 2^{n-1} - R_f/2$. *Бент-функциями* называются функции от четного числа переменных с максимальной возможной нелинейностью. Они были впервые описаны в [1]. Подробную информацию об этом классе функций можно найти в [2, 3]. Булевы функции f и g от n переменных *аффинно эквивалентны*, если для всех x выполнено $g(x) = f(Ax+b)$, где A — невырожденная матрица размера $n \times n$, а b — вектор длины n.

В практических целях также часто требуется чтобы функция была *сбалансированной* — принимала значения 0 и 1 на одном и том же числе аргументов. Но максимальное значение нелинейности сбалансированных функций неизвестно начиная уже с восьми переменных. Лучшие оценки получаются как следствие конкретных конструкций сбалансированных функций.

Конструкция, описанная Доббертином в [4], основана на модификации нормальных бент функций — функций от $2n$ переменных, постоянных на некотором аффинном подпространстве L размерности n. Суть конструкции заключается в замене значений

[1]Работа выполнена в рамках государственного задания ИМ СО РАН (проект № 0314–2019–0017) при поддержке Российского Фонда Фундаментальных Исследований (проект 20–31–70043) и лаборатории криптографии JetBrains Research.

бент-функции на подпространстве L значениями сбалансированной функции θ от n переменных. При этом спектральный радиус получившейся сбалансированной функции Θ равен $R_\Theta = 2^n + R_\theta$, а её нелинейность, соответственно $N_\Theta = 2^{2n-1} - 2^{n-1} - R_\theta/2$. Также в [4] была сформулирована не опровергнутая до сих пор гипотеза о несуществовании сбалансированных функций с нелинейностью выше, чем можно получить при помощи этой конструкции.

В работе рассмотрено обобщение конструкции Доббертина, использующее бент-функции с близкими к нормальности свойствами, а именно бент-функции от $2n$ переменных, принимающие постоянное значение на нескольких сдвигах некоторого подпространства L размерности $n-k$, здесь $0 \leqslant k \leqslant n-2$. Так как аффинная эквивалентность сохраняет нелинейность и сбалансированность, мы можем без ограничения общности рассматривать такие бент-функции в виде $f : \mathbb{F}_2^{n-k} \times \mathbb{F}_2^{n+k} \to \mathbb{F}_2$, для которой существуют подмножества $I_0, I_1 \subset \mathbb{F}_2^{n+k}$, мощности $|I_0| = 2^{2k-1} + 2^{k-1}$, $|I_1| = 2^{2k-1} - 2^{k-1}$, для которых справедливо:

$$f(x,y) \equiv 0, \text{ при } y \in I_0$$

$$f(x,y) \equiv 1, \text{ при } y \in I_1$$

Такое представление прямо связанно с конструкцией вида $\widetilde{f} \oplus Ind_{L^\perp}$, подробную информацию о которой можно найти в [5, 6, 7]. Здесь $\widetilde{f}$ — дуальная к f функция, см.[3].

При помощи бент функции такого вида и набора θ_y, $y \in I_0 \cup I_1$ сбалансированных функций от $n-k$ переменных строится обобщающая конструкцию Доббертина функция Θ:

$$\Theta(x,y) = \begin{cases} \theta_y(x), & \text{при } y \in I_0 \cup I_1 \\ f(x,y), & \text{иначе.} \end{cases} \tag{1}$$

При $k = 0$ описанная конструкция полностью совпадает с конструкцией Доббертина. При $k = 1$ она также эквивалентна конструкции Доббертина. Для функции Θ выполненны:

Теорема 1. Функция Θ вида (1) является сбалансированной функцией и её коэффициенты Уолша-Адамара вычисляются по формуле

$$W_\Theta(a,b) = \begin{cases} W_f(a,b) + \sum\limits_{y \in I_0 \cup I_1} (-1)^{\langle b,y \rangle} W_{\theta_i}(a), & \text{если } a \neq 0 \\ 0, & \text{иначе} \end{cases}.$$

Следствие 1. Спектральный радиус Θ не превосходит $2^n + \sum_{y \in I_0 \cup I_1} R_{\theta_y}$, причем всегда можно выбрать θ_y, при которых оценка достигаеся.

Теорема 2. Пусть θ — сбалансированная функция $n-k$ переменных, $\theta_y = \theta$ при $y \in I_0$, и $\theta_y = \theta \oplus 1$ при $y \in I_1$. Тогда

$$R_\Theta = 2^n + 2^k R_\theta.$$

Получившееся R_Θ зависит от R_θ, k и n. Несмотря на то, что θ является функцией от $n-k$ переменных, наилучший результат достигается при $k = 0$, то есть в случае, описанном Доббертином.

ЛИТЕРАТУРА

1. Rothaus O. On «bent» functions // J. Combin. Theory, Ser. A. V. 20. No. 3. pp. 300–305. 1976.

2. Логачев О. А., Сальников А. А., Смышляев С. В., Ященко В. В. Булевы функции в теории кодирования и криптологии. 2-е изд. М.: МЦНМО, 584 с. 2012.
3. Tokareva N. N. Bent Functions, Results and Applications to Cryptography. Acad. Press. Elsevier, 2015.
4. Dobbertin. H. Construction of bent functions and balanced Boolean functions with high nonlinearity. // LNCS V. 1008, pp. 61–74. Springer-Verlag, 1994.
5. Kolomeec N. On properties of a bent function secondary construction // Тезисы BFA'2020 (28 сентября – 2 октября 2020), будут доступны на сайте https://boolean.w.uib.no/bfa-2020
6. Коломеец Н. А., О некоторых свойствах конструкции бент-функций с помощью подпространств произвольной размерности, ПДМ. Приложение, No 11, с. 41–43, 2018.
7. Carlet C. Two new classes of bent functions // LNCS. V. 765. pp. 77—101. 1994.

Sutormin I.A. **AN ESTIMATION OF THE NONLINEARITY OF BALANCED BOOLEAN FUNCTIONS GENERATED BY GENERALIZED DOBBERTIN'S CONSTRUCTION.** A generalization of Dobbertin's construction for highly nonlinear balanced boolean functions is proposed. Their Walsh-Hadamard spectrum was studied and estimates of the spectral radius of the resulting functions were obtained. An exact upper bound for the spectral radius is proved and a method is proposed for constructing a balanced function of $2n$ variables using a balanced θ of $n-k$ variables with a spectral radius of $2^n + 2^k R_\theta$, where R_θ is the spectral radius of θ.
Keywords: *boolean functions, bent functions, balancedness, nonlinearity, spectral radius*

СУТОРМИН Иван Александрович — м.н.с. Института математики им. С.Л.Соболева СО РАН, студент Новосибирского государственного университета, Новосибирск. E-mail: **ivan.sutormin@gmail.com**

УДК 519.7

СВЯЗЬ МЕЖДУ КВАТЕРНАРНЫМИ И КОМПОНЕНТНЫМИ БУЛЕВЫМИ БЕНТ-ФУНКЦИЯМИ[1]

А. С. Шапоренко

В работе исследуются кватернарные бент-функции. Функция $g : \mathbb{Z}_4^n \to \mathbb{Z}_4$ называется кватернарной функцией от n переменных. В работе доказано, что свойство кватернарной функции $g(x + 2y) = a(x, y) + 2b(x, y)$ быть бент напрямую не зависит от того, является ли функции b и $a \oplus b$ булевыми бент-функциями. Получено количество кватернарных бент-функций от одной и двух переменных с описанием свойств булевых функций b и $a \oplus b$. Представлены простые конструкции кватернарных бент-функций от любого числа переменных.

Ключевые слова: *кватернарные функции, булевы функции, бент-функции*

Пусть $\langle x, y\rangle$ обозначает скалярное произведение двоичных векторов по модулю 2 (обозначим $\oplus$), а $x.y$ – скалярное произведение векторов по модулю 4.

Функция $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ называется *булевой функцией* от n переменных. *Преобразованием Уолша–Адамара булевой функции* f от n переменных называется целочисленная функция $W_f(x)$, заданная на множестве $\mathbb{Z}_2^n$ равенством

$$W_f(x) = \sum_{y \in \mathbb{Z}_2^n} (-1)^{\langle x,y\rangle \oplus f(y)}.$$

Булева функция f от n (четное) переменных называется *бент-функцией*, если $|W_f(x)| = 2^{n/2}$ для любого $x \in \mathbb{Z}_2^n$.

Шифры, в которых используются бент-функции, более устойчивы к *линейному криптоанализу* [1], потому что бент-функции крайне плохо аппроксимируются аффинными функциями. Бент-функции использовались в дизайне блочного шифра *CAST* как координатные функции S–блоков [2], а также для построения регистра сдвига с нелинейной обратной связью в поточном шифре *Grain* [3]. Также бент-функции связаны с некоторыми объектами теории кодирования, например, с *кодами Рида-Маллера* [4].

Функция $g : \mathbb{Z}_4^n \to \mathbb{Z}_4$ называется *кватернарной функцией* от n переменных [5]. *Преобразование Уолша–Адамара кватернарной функции* g определяется следующим образом:

$$W_g(x) = \sum_{y \in \mathbb{Z}_4^n} i^{x.y+g(y)},$$

где '+' означает сложение по модулю 4.

Кватернарная функция g от n переменных называется *бент-функцией*, если $|W_g(x)| = 4^{n/2}$ для любого $x \in \mathbb{Z}_4^n$.

Целью данной работы является изучения связи свойств быть бент кватернарных и булевых функций. Эта задача была впервые поставлена в работе [6] (см. также [7]).

Каждая кватернарная функция g от n переменных может быть представлена для любых $x, y \in \mathbb{Z}_2^n$ следующим образом:

$$g(x + 2y) = a(x, y) + 2b(x, y),$$

[1]Работа выполнена в рамках государственного задания ИМ СО РАН (проект № 0314-2019-0017) при поддержке Российского Фонда Фундаментальных Исследований (проект 18-07-01394) и лаборатории криптографии JetBrains Research.

где сложение производится по модулю 4, а функции a и b – это *компонентные* булевы функции от $2n$ переменных.

Утверждение 1. Для любой кватернарной функции $g(x+2y) = a(x,y)+2b(x,y)$ от одной переменной, где $x, y \in \mathbb{Z}_2$, справедливо, что g – кватернарная бент-функция тогда и только тогда, когда $b(x,y)$ – бент-функция и $a(x,y)$ равна $0, 1, x$ или $x \oplus 1$. Кроме того, если g – кватернарная бент-функция, тогда b и $a \oplus b$ – булевы бент-функции.

Компьютерные вычисления показали, что количество кватернарных бент-функций от одной переменной равно 32.

Количество кватернарных бент-функций при $n = 2$ равно 200704. Среди них 98304 функций таких, что ни одна из булевых функций a, b и $a \oplus b$ не является бент-функцией, но при этом для 3072 из них a линейная. Существуют 36864 функции таких, что b и $a \oplus b$ – бент-функции, при этом для 33792 из них функция a нелинейная, а для 2304 и 768 a является линейной функцией или константой соответственно. Количество кватернарных функций, для которых каждая из функций a, b и $a \oplus b$ – бент-функция, равно 16384. Для оставшихся 49152 функций a является бент–функцией, а b и $a \oplus b$ нелинейные булевы функции.

Теорема 1. Пусть $g(x+2y) = a(x,y)+2b(x,y)$ – кватернарная бент-функция, где $x, y \in \mathbb{Z}_2^n$ и a, b – булевы функции от $2n$ переменных. Тогда b и $a \oplus b$ – нелинейные функции при любом числ переменных $n \geqslant 1$.

Следующие два утверждения показывают, что между свойствами быть бент кватернарной функции g и ее компонентных булевых функций b и $a \oplus b$ нет прямой связи.

Утверждение 2. Для любого $n \geqslant 2$ существует кватернарная бент-функция $g(x+2y) = a(x,y)+2b(x,y)$ от n переменных, где b и $a \oplus b$ не являются бент-функциями от $2n$ переменных.

Утверждение 3. Для любого n существует кватернарная функция $g(x+2y) = a(x,y)+2b(x,y)$ от n переменных, которая не является бент-функцией, когда b и $a \oplus b$ – булевы бент функции от $2n$ переменных.

Далее представим две простые конструкции для кватернарных бент-функций от любого числа переменных.

Утверждение 4. Кватернарная функция от n переменных

$$g(x_1 + 2x_{n+1}, ..., x_n + 2x_{2n}) = \sum_{i=1}^{n} 2x_i x_{i+n} + cx_j,$$

где $c \in \mathbb{Z}_2$, $j \in \{1, ..., n\}$ и '+' – сложение по модулю 4, является бент-функцией при любом n. Заметим, что при этом

$$b(x_1, ..., x_{2n}) = \bigoplus_{i=1}^{n} x_i x_{i+n},$$

$$a(x_1, ..., x_{2n}) \oplus b(x_1, ..., x_{2n}) = \bigoplus_{i=1}^{n} x_i x_{i+n} \oplus cx_j$$

– бент-функции от $2n$ переменных.

Утверждение 5. Пусть $g(x+2y) = a(x,y)+2b(x,y)$, где $x, y \in \mathbb{Z}_2^n$ и a и b — булевы функции от $2n$ переменных, является бент-функцией, тогда функция $g'(x+2y) = 3a(x,y)+2b(x,y)$ также является кватернарной бент-функцией от $n \geqslant 1$ переменных.

Отметим, что утверждение верно и в обратную сторону.

ЛИТЕРАТУРА

1. *Matsui M.* Linear Cryptanalysis Method for DES cipher //Advances in Cryptology – Eurocrypt 1993, Springer-Verlog, Berlin, pp. 386–397.
2. *Adams C.* Constructing symmetric ciphers using the CAST design procedure //Proc. Design, Codes, and Cryptography, vol. 12, no. 3, pp. 283–316, 1997.
3. *Hell M., Johansson T., Maximov A., and Meier W.* A stream cipher proposal: Grain-128 //IEEE International Symposium on Information Theory, pp. 1614–1618, 2006.
4. *Tokareva N.* Bent functions: results and applications to cryptography //Acad. Press. Elsevier, 2015.
5. *Kumar P. V., Scholtz R. A., Welch L. R.* Generalized bent functions and their properties //Journal of Combinatorial Theory, vol. 40, no. 1, pp. 90—107, 1985.
6. *Solé P., Tokareva N.* Connections Between Quaternary and Binary Bent Functions //Cryptology ePrint Archive, Report 2009/544, available at `http://eprint.iacr.org/`.
7. *Solé P., Tokareva N.* On Quaternary and Binary Bent Functions. //Prikl. Diskr. Mat., 2009, supplement no 1, pp. 16–18. Available at `www.mathnet.ru`.

Shaporenko A. S. **CONNECTIONS BETWEEN QUATERNARY AND COMPONENT BOOLEAN BENT FUNCTIONS.** This work is about quaternary bent functions. Function $g : \mathbb{Z}_4^n \to \mathbb{Z}_4$ is called quaternary on n variables. It was proven that bentness of a quaternary function $g(x + 2y) = a(x, y) + 2b(x, y)$ doesn't directly depend on the bentness of Boolean functions b and $a \oplus b$. The number of quaternary bent functions in one and two variables is obtained with a description of properties of Boolean functions b and $a \oplus b$. Two simple constructions of quaternary bent functions in any number of variables are presented.
Keywords: *quaternary functions, boolean functions, bent function.*

Шапоренко Александр Сергеевич — м.н.с Института математики им. С. Л.Соболева СО РАН, студент Новосибирского государственного университета, исследователь лаборатории криптографии JetBrains Research, Новосибирск. E-mail: **shaporenko.alexandr@gmail.com**

УДК 519.7

Разработка методов анализа блокчейн сетей

Д. А. Бадер
Новосибирский государственный университет

В данной работе рассматриваются транзакции блокчен сети Ethereum. Идея исследования состоит в том, чтобы провести временной анализ транзакций блокчейн сети Ethereum и оценить, как изменялось поведение аккаунтов со временем. Планируется провести анализ контрактных аккаунтов, их создания и назначения, а также провести анализ типов Ethereum-токенов. Это поможет понять как и для чего используют криптовалюту Ethereum.

Блокчейн - выстроенная по определенным правилам непрерывная последовательная цепочка (связный список) блоков, содержащих информацию. Основная задача технологии блокчейн - доверительная передача собственности на цифровые активы в недоверенной среде без посредников. *Транзакция* - единственный способ изменить состояние данных. *Блок* - структура данных, позволяющая хранить список транзакций. *Криптовалюта* - это реализация блокчейн. *Миксер* - сервис анонимизации, который усложняет или делает практически невозможным отслеживание транзакций в системе блокчейн.

В работе проведен временной анализ по количеству транзакций, отправленных на адрес, и количество транзакций, отправленных с него. Был составлен график, показывающий количество адресов с определенным набором исходящих и входящих транзакций. На основе этого графика был замечен аномальный рост числа адресов с двумя отправителями и тремя получателями с февраля 2017 по апрель 2018. Был построен граф этих адресов, что показало, что эти адреса связаны в большой миксер.

В работе был проведен временной анализ токенов ERC-20, ERC-223, ERC-721, ERC-827 на предмет того, какие токены выпускаются и как они используются. В результате видно, что интерес к созданию новых токенов не спадает. То же касается количества переводов токенов, что показывало их активное использование.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

---

[1] Buterin, V.: Ethereum: a next generation smart contract and decentralized applicationplatform (2013). URL: https://github.com/ethereum/wiki/wiki/White-Paper

Научный руководитель – канд. физ.-мат. наук Н. Н. Токарева, П.А. Сазонова

УДК 519.7

Легковесные шифры типа Лая-Мэсси

А. А. Белоусова
Новосибирский государственный университет

В данной работе рассматриваются блочные итеративные шифры, основанные на сети Фейстеля и на альтернативной схеме – схеме Лая-Мэсси. Идея исследования состоит в том, чтобы рассмотреть шифр Simon $32/64$, основаный на сети Фейстеля, и сравнить его криптографические свойства со свойствами адаптации схемы Лая-Мэсси на место сети Фейстеля. Результаты дифференциального криптоанализа шифра Simon были взяты из работы [2], где было получено, что для Simon $32/64$ максимальная вероятность дифференциала после прохождения 12 раундов составляет 2^{-36}.

Один раунд схемы Лая-Мэсси в её оригинальном виде записывается как $(y_L, y_R) = (x_L \oplus F(x_L \oplus x_R), x_L \oplus F(x_L \oplus x_R))$ и в данном случае есть существенный недостаток: для любого входа (x_L, x_R) выполняется соотношение $x_L \oplus x_R = y_L \oplus y_R$, где (y_L, y_R) это выход раунда.

В работе [1] сказано, что для того чтобы убрать описаный выше недостаток к схеме необходимо добавить перестановку-ортоморфизм σ.

Определение 1. Пусть $\sigma\colon \mathbb{Z}_n \to \mathbb{Z}_n$ перестановка на $\mathbb{Z}_n$, σ называется ортоморфизмом $\mathbb{Z}_n$, если $\sigma + I$ так же является перестановкой на $\mathbb{Z}_n$, где I - тождественная перестановка.

Тогда один раунд схемы будет записан как $(y_L, y_R) = (\sigma(x_L \oplus F(x_L \oplus x_R)), x_R \oplus F(x_L \oplus x_R))$, а разница текстов будет записана как $y_L \oplus y_R = (\sigma(x_L \oplus F(x_L \oplus x_R)) \oplus (x_L \oplus F(x_L \oplus x_R))) \oplus (x_L \oplus x_R)$. Для сравнения шифров был проведён дифференциальный криптоанализ оригинальной схемы и схемы с добавлением ортоморфизма.

Утверждение 1. После 12 раундов максимальная вероятность дифференциала для модернизированного шифра Simon$32/64$ без добавления ортоморфизма составляет 2^{-24}, а с добавлением ортоморфизма $\geqslant 2^{-63}$.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

[1] Vaudenay S. On the Lai-Massey Scheme. Ecole Normale Superieure,1999.

[2] Abed F., List E., Lucks S., Wenzel J. Differential and Linear Cryptanalysis of Reduced-Round Simon. Bauhaus-University Weimar, Germany, 2013.

Научный руководитель – канд. физ.-мат. наук Н.Н. Токарева

УДК 519.7

Анализ гаммы, порожденной фильтрующим генератором

Т. А. Бонич
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

Для построения поточных шифров часто используются регистры с обратной связью. Наибольшее распространение получили регистры сдвига с линейными обратными связями (РСЛОС). РСЛОС состоит из двух частей: бинарный вектор $x = (x_{n-1}, \dots x_0)$ длины n и определенная на нем функция обратной связи $f : (x_{n-1}, \dots, x_0) \to \{0, 1\}$, где f – булева функция от n переменных. Фильтрующий генератор состоит из одного регистра сдвига с линейной обратной связью длины n, для изменения состояний использует примитивный многочлен. Булева функция $h(x_{n-1}, \dots, x_0)$ будет генерировать последовательность γ. Работа генератора представлена, например, в [1]. Пусть $\gamma = (y_1 y_2 \dots y_{2^n-1})$, где $h(x_{n-1}, \dots, x_0) = y_1, h(x_{n-2}, \dots, x_0, f(x_{n-1}, \dots, x_0)) = y_2$, и т.д. Так как нулевое состояние не используется, тогда количество всех возможных состояний генератора равно $2^n - 1$. Тогда, булева функция может генерировать γ с периодом от 1 до $2^n - 1$. В данной работе изучено, как выбор булевой функции h влияет на периодические свойства генерируемой γ. А именно, определено количество всех булевых функций h, которые порождают последовательность не максимального периода ($< 2^n - 1$). Такие функции будем называть *неподходящими*.

Теорема 1. *Пусть* $2^n - 1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_s}$, *где* p_i – *различные простые целые числа,* α_i *- положительное целое число,* s – *количество чисел, участвующих в разложении. Тогда количество неподходящих булевых функций от n переменных для фильтрующего генератора равно*

$$2 \cdot \sum_{\beta \in \mathbb{Z}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \dots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \dots p_s^{\alpha_s - \beta_s}}),$$

где $\beta = (\beta_1, \dots, \beta_s)$.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

[1] Н. Н. Токарева. Симметричная криптография. Краткий курс. – Новосибирский государственный университет. – 2012.

Научный руководитель – канд. физ.-мат. наук, доц. Н.Н.Токарева

УДК 519.7

Тесты для SAT-решателей, основанные на криптографических задачах

А. Е. Доронин
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

В настоящее время в криптографии SAT-решатели используются для проведения криптоанализа семейства шифров Trivium в работе [1] и некоторых поточных шифров в работе [2]. В данной работе предлагается их использование в задачах поиска криптографических булевых функций и проверки эквивалентности двух функций. Для получения набора булевых формул были использованы следующие понятия и свойства:

Векторная булева функция $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ является *взаимно-однозначной*, если выполняется одно из следующих условий:

$\forall x_1 \in \mathbb{Z}_2^n \; \forall x_2 \in \mathbb{Z}_2^n : x_1 \neq x_2 \to F(x_1) \neq F(x_2)$,

$\forall y \in \mathbb{Z}_2^n \; \exists! x \in \mathbb{Z}_2^n : F(x) = y$.

Векторная булева функция $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ является *дифференциально δ-равномерной*, если для любых $a \neq 0$, b уравнение $F(x) \oplus F(x \oplus a) = b$ имеет не более δ решений.

Векторные булевы функции F и G называются *EA-эквивалентными*, если выполняется следующее: $G = B \circ F \circ A + C$, где A, B и C - аффинные функции.

Данные понятия представляются в виде КНФ и подаются на вход SAT-решателя. В результате его работы происходит означивание переменных таким образом, чтобы формулы были истинными. Полученный набор формул также можно использовать для тестирования работы новых SAT-решателей, созданных для решения криптографических задач.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

[1] О. С. Заикин, И. В. Отпущенников, А. А. Семёнов, “Оценки стойкости шифров семейства Trivium к криптоанализу на основе алгоритмов решения проблемы булевой выполнимости”, ПДМ. Приложение, 2016, № 9, 46–48

[2] А. С. Игнатьев, А. А. Семенов, Д. В. Беспалов, О. С. Заикин, “Гибридный подход (SAT+ROBDD) в задачах криптоанализа поточных систем шифрования”, ПДМ, 2009, приложение № 1, 19–20

Научный руководитель – канд. физ.-мат. наук К. В. Калгин,
канд. физ.-мат. наук Н. Н. Токарева

УДК 004.056.55

Криптоанализ базовой версии криптографической системы с открытым ключом, основанной на сложности решения системы полиномиальных уравнений в целых числах

Е. В. Завалишина
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

В 2016 году the National Institute of Standards and Technology представил доклад под названием Report on Post-Quantum Cryptography, в котором полагает, что пришло время подготовиться к переходу на квантово-устойчивую криптографию, так как некоторые задачи, лежащие в основе использующихся на практике криптографических алгоритмов, могут быть решены квантовыми компьютерами.

В связи с этим автором настоящей работы и соавторами была предпринята попытка создать новый алгоритм шифрования данных с открытым ключом, основанный на решении системы однородных полиномиальных уравнений в целых числах, описанный в статье [1].

Данная работа посвящена криптоанализу описанной системы. Автором работы был разработан алгоритм атаки на основе подобранного открытого текста, который позволяет получить набор матриц, которые могут использоваться в качестве закрытого ключа.

Так как набор полиномов, использующийся в качестве открытого ключа, имеет строго определенный вид, следовательно, возможно выразить коэффициенты полиномов открытого ключа через элементы матриц. Так как коэффициенты известны, можно составить систему уравнений, решение которой даст набор искомых матриц.

На основе данного исследования проведена оценка целесообразности усложнения системы и пути ее реализации.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

[1] Волков, Е., Баранов А., Завалишина Е. Криптографическая система с открытым ключом // Second Conference on Software Engineering and Information Management (SEIM-2017), 2017. С. 41-44.

Научный руководитель - канд. физ.-мат. наук, доц. Токарева Н. Н.

УДК 519.7

О числе взаимно однозначных векторных булевых функций специального вида

М. М. Запольский
Новосибирский государственный университет

S-блоки [1] являются частью блочных шифров. С точки зрения математики S-блок – это биективная векторная булева функция, обладающая рядом свойств, обеспечивающих криптостойкость. На практике поиск подходящих функций перебором даже при малом числе переменных не представляется возможным. Мы будем изучать некоторые классы векторных булевых функций, и искать число биекций в них.

Пусть $\pi \in S_n$ – произвольная перестановка. Рассмотрим бинарный вектор $x \in \mathbb{F}_2^n$, $x = (x_1, \ldots, x_n)$, обозначим $\pi(x) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. Пусть f – булева функция n переменных, построим векторную булеву функцию $F_\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ следующим образом:

$$F_\pi(x) = (f(x), f(\pi(x)), f(\pi^2(x)), \ldots, f(\pi^{n-1}(x))).$$

Также введем множество $\Delta_{\pi,n}$ всех таких векторных булевых функций.
Пусть $\rho(x) = (x_n, x_1, x_2, \ldots, x_{n-1})$, – циклический сдвиг.

Предложение 1. *Пусть $\pi \in S_n$, $F_\pi \in \Delta_{\pi,n}$. Для любого $x \in \mathbb{F}_2^n$ и $k \in \mathbb{Z}$ справедливо соотношение: $F_\pi(\pi^k(x)) = \rho^{-k}(F_\pi(x))$.*

Введем действие π на множестве $\mathbb{F}_2^n$ следующим образом: $x \circ \pi = \pi(x)$. Данное действие разбивает $\mathbb{F}_2^n$ на орбиты относительно π. За $O_\pi(x)$ обозначим орбиту, порожденную $x \in \mathbb{F}_2^n$. Обозначим через $\Theta_{\pi,n}$ множество всех орбит относительно π. Из предложения 1 следует:

Предложение 2. *Пусть $\pi \in S_n$, $F_\pi \in \Delta_{\pi,n}$. Для любой орбиты $o \in \Theta_{\pi,n}$ существует $g \in \Theta_{\rho,n}$ такая, что образ множества o под действием F_π лежит в g.*

Введем отображение $\Psi_{F_\pi,n} : \Theta_{\pi,n} \to \Theta_{\rho,n}$ так: $\Psi_{F_\pi,n}(O_\pi(x)) = O_\rho(F_\pi(x))$. Данное отображение корректно определено в силу предложения 2.

Предложение 3. *$F_\pi \in \Delta_{\pi,n}$ взаимно однозначна тогда и только тогда $\Psi_{F_\pi,n}$ взаимно однозначная функция. Если $\Psi_{F_\pi,n}$ биекция, тогда для каждой орбиты $o \in \Theta_{\pi,n}$ выполнено: $|\Psi_{F_\pi,n}(o)| = |o|$.*

Обозначим через $M^k_{\pi,n}$ множество орбит $o \in \Theta_{\pi,n}$ таких, что $|o| = k$. Будем считать $M^k_{\rho,n} = M^k_n$. Заметим, что выполнено: $2^k = \sum_{\ell:\ell|k} \ell \cdot |M^\ell_n|$

Теорема 1. *Если $|M^k_{\pi,n}| = |M^k_n|$ для всех k, тогда число взаимно однозначных функций из $\Delta_{\pi,n}$ равно $\prod_{k:k|n} |M^k_n|! \cdot k^{|M^k_n|}$, иначе оно пусто.*

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

---

[1] Токарева Н. Н. Симметричная криптография. Краткий курс // Учебное пособие: Новосибирский государственный университет, 2012.

Научный руководитель – канд. физ.-мат. наук. Н. Н. Токарева

УДК 519.7

Криптографические свойства S-блока, построенного на основе булевой функции и перестановки

Д. А. Зюбина
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

S-блоки играют решающую роль в обеспечении стойкости блочных шифров относительно разных типов атак. *S-блок* - это отображение из множества двоичных векторов длины n в себя. Проектироваться S-блоки должны наиболее тщательно, так как стойкость всего шифра существенно зависит от их криптографических характеристик. В данной работе S-блок представлен в виде векторной булевой функции $F_\pi(x) = \left(f(x), f\left(\pi(x)\right), f(\pi^2(x)), \dots, f(\pi^{n-1}(x))\right)$, где f - булева функция от n переменных, π - перестановка n элементов. Были изучены криптографические свойства F_π (такие как *алгебраическая степень*, *нелинейность, дифференциальная δ-равномерность* и *уравновешенность*) в зависимости от свойств f и перестановки π (например, перестановка-беспорядок[1]) при малых значениях n. Векторная булева функция F называется дифференциально δ-равномерной, если при любом векторе $a \neq 0$ и произвольном векторе b уравнение $F(x) \oplus F(x \oplus a) = b$ имеет не более δ решений, где δ - целое число. Минимальное возможное значение δ равно 2. Пусть A_4^1 множество перестановок-беспорядков для 4 элементов, содержащее три пары перестановок таких, что $\pi_{2_i}^{-1} = \pi_{1_i}, i = 1,2,3$: $\pi_{1_1} = (2\ 3\ 4\ 1), \pi_{2_1} = (4\ 1\ 2\ 3), \pi_{1_2} = (2\ 4\ 1\ 3), \pi_{2_2} = (3\ 1\ 4\ 2), \pi_{1_3} = (3\ 4\ 2\ 1), \pi_{2_3} = (4\ 3\ 1\ 2)$.

Утверждение 1. Для n = 2,3 существует булева функция f от n переменных и перестановка-беспорядок $\pi \in S_n$ такая, что $\delta_{F_\pi} = 2$. Для всякой перестановки $\pi \in A_4^1$ существует булева функция f от 4 переменных такая, что $\delta_{F_\pi} = 2$.

Данные результаты будут использованы для построения S-блока на основе булевой функции и перестановки с необходимыми криптографическими свойствами. Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

---

1. Р.Стенли. Перечислительная комбинаторика. — М.: Мир, 1990.

Научный руководитель – к.ф.-м.н. Н.Н.Токарева

УДК 519.7

Анализ гаммы, порождаемой комбинирующим генератором

М. А. Панферов
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

Регистры сдвига с линейной обратной связью используются для построения генераторов в поточных шифрах. Комбинирующие генераторы состоят из несколько регистров сдвига с линейной обратной связью, причем каждый регистр имеет свою длину n_i и использует свой примитивный многочлен для изменения состояний. Булева функция $h(X_1, \ldots, X_m)$, где X_i – битовая строка регистра i, будет генерировать псевдослучайную последовательность *гамма*. Заполнение векторов $X_1, \ldots, X_m$ конкретными значениями будем называть состоянием регистров. Состояние нулевое, если все $X_i = (0, \ldots, 0)$. Работа комбинирующего генератора подробнее описана в [1]. Так как мы не используем нулевое состояние в каждом регистре, то общее количество состояний регистров не превосходит $(2^{n_1} - 1)(2^{n_2} - 1) \ldots (2^{n_m} - 1)$. При этом максимум достигается при $(n_i, n_j) = 1$, где $i, j = 1, \ldots, m$, $i \neq j$. В данной работе исследовалось, как выбор функции h влияет на периодические свойства генерируемой гаммы. А именно, определено количество булевых функций h, которые порождают последовательность с периодом меньше, чем $(2^{n_1} - 1)(2^{n_2} - 1) \ldots (2^{n_m} - 1)$. Такие функции будем называть *неподходящими*.

Теорема 1. *Пусть m – количество регистров с длинами $n_1, \ldots, n_m$. И $(2^{n_1} - 1)(2^{n_2} - 1) \ldots (2^{n_m} - 1) = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_s^{\alpha_s}$, где p_i различные простые числа, $\alpha_i > 0$. Тогда количество неподходящих булевых функций от n переменных для комбинирующего генератора равно*

$$2 \cdot \sum_{\beta \in \mathbb{F}_2^s, \beta \neq 0} ((-1)^{\beta_1 + \cdots + \beta_s + 1} 2^{p_1^{\alpha_1 - \beta_1} \ldots p_s^{\alpha_s - \beta_s}}),$$

где $\beta = (\beta_1, \ldots, \beta_s)$.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

[1] Н. Н. Токарева. Симметричная криптография. Краткий курс. – Новосибирский государственный университет. – 2012.

Научный руководитель – к.ф.-м.н. Н. Н. Токарева

УДК 519.7

О декомпозиции векторных булевых функций

Г. М. Пинтус
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research, г. Новосибирск

Векторной булевой (n, m)*-функцией* называется произвольное отображение вида $\mathbb{F}_2^n \to \mathbb{F}_2^m$. Векторная булева (n, m)-функция F задается набором из m координатных булевых функций от n переменных правилом $F(x) = (f_1(x), f_2(x), ..., f_m(x)), x \in \mathbb{F}_2^n$. Нетривиальные линейные комбинации функций $\{f_i\}_{i=1}^m$ называются *компонентными функциями*. Минимальная из алгебраических степеней компонентных функций называется алгебраической *степенью* функции F и обозначается через $\deg(F)$.

Задача нахождения декомпозиции векторной булевой (n, n)-функции F состоит в поиске двух векторных булевых (n, n)-функций G, H, таких что $\max\{\deg(G), \deg(H)\} < \deg(F)$ и $F(x) = G(H(x))$ для всех $x \in \mathbb{F}_2^n$. Векторную булеву (n, n)-функцию F степени $d > 2$, допускающую такую декомпозицию, будем называть *разложимой*. Решение данной задачи применительно к векторным булевым функциям, описывающим нелинейные преобразования раундовой функции симметричного блочного шифра, имеет прямое отношение к защите от атак по сторонним каналам [1].

Две векторные (n, n)-функции F и G называются *расширенно аффинно эквивалентными* (EA-эквивалентными), если существуют две аффинные (n, n)-подстановки A, B и аффинная (n, n)-функция C, такие что $G(x) = (B \circ F \circ A)(x) + C(x), x \in \mathbb{F}_2^n$.

Утверждение 1. *Пусть (n, n)-функция F степени $d > 2$ разложима. Тогда (n, n)-функция $F' = A_2 \circ F \circ A_1$, где A_1, A_2 — произвольные аффинные (n, n)-подстановки, также будет являться разложимой.*

Если F представима в виде композиции двух (n, n)-функций G, H степени меньше d, таких что функция H обратима, и для функции H^{-1} справедливо условие $\deg\left(H^{-1}\right) \leqslant \max\{\deg(G), \deg(H)\}$, то (n, n)-функция $F'' = F + A_0$ будет разложимой для любой аффинной (n, n)-функции A_0 .

[1] Bilgin B., Nikova S., Nikov V., Rijmen V., Tokareva N., Vitkup V., Threshold implementations of small S-boxes, Cryptogr. Commun., **7**(1), 33–33 (2015).

Научный руководитель – к.ф.-м.н., доц. Н. Н. Токарева, А. В. Куценко

УДК 004.056.5

Применение SAT решателей в криптоанализе

Д.А. Софронова
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

В основе одного из методов анализа шифров лежит использование SAT-решателей. Для этого исходный шифр или хэш-функция записываются в виде логической формулы, истинность которой предстоит установить SAT-решателю.SAT-задача – задача определения выполнимости логической формулы[1].SAT-решатель – программа, которая ищет означивание переменных, на котором формула истинна.

Существуют два проекта для описания криптографических шифров – Transalg[3] и Grainofsalt[2]. Первый имеет заметный недостаток – отсутствие документации, затрудняющее использование. Второй подходит только для описания шифров, основанных на регистрах сдвига, хотя хорошо оптимизирован и удобен в работе.

Цель данной работы – разработка легковесного, гибкого, свободно расширяемого программного комплекса для описания шифров, хэш-функций и криптографических задач, позволяющего получать как КНФ, так и реализацию самого алгоритма для проверки правильности реализации и тестирования криптостойкости другими средствами (NIST, Dieharder).

Реализован простой предметно-ориентированный язык описания шифров – основа проекта. Ядро программы насчитывает 250 строк кода на C++, что обеспечивает легкую расширяемость функционала.ОписанышифрыA5/1, Grain, Crypto1, Bivium, Trivium, Hitag2. Были проведены атаки угадай-и-вычисли на шифр A5/1.

Работа выполнена при поддержке лаборатории криптографии JetBrainsResearch.

[1] Armin Biere, Marijn Heule, Hans van Maaren, Toby Walsh, “HANDBOOK OF SATISFIABILITY, Frontiers in Artificial Intelligence and Applications”. IOS Press, 2009

[2] Mate Soos, “Grain of Salt — An Automated Way to Test Stream Ciphers through SAT Solvers**”**, Workshop on Tools for Cryptanalysis, Royal Holloway, University of London, 2010.

[3] И. В. Отпущенников, А. А. Семёнов, “Технология трансляции комбинаторных проблем в булевы уравнения”, ПДМ, 2011

Научный руководитель – канд.физ.-мат.наук К.В. Калгин

УДК 519.7

Алгоритм меж-блокчейн взаимодействия для сценария залогового удержания

А. Д. Сычев
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

С ростом различных реализаций блокчейнов - децентрализованных баз данных транзакций, в которых транзакции хранятся в структуре данных, называемой блоки, - появляется необходимость решения задачи их взаимодействия. С технической точки зрения можно выделить 3 типа взаимодействия блокчейнов: Нотариальные схемы, Боковые цепи и Хэш-блокировка. Рассмотрим Хэш-блокировку - алгоритм, описывающий действия в цепочке A и цепочке B, которые имеют один и тот же триггер, как правило, обнаружение прообраза конкретного хэша. Более подробное описание можно найти в статье В. Бутерина [1].

На данный момент существует алгоритм "Атомарного обмена"[2] валютами в двух различных сетях, основанный на Хэш-блокировке. Это значит, что в результате сделки в обеих цепях валюта либо дойдет до получателя, либо вернется обратно к владельцу.

Основным результатом данной работы является создание алгоритма для сценария залогового удержания, использующего идею "Атомарного обмена" с некоторыми изменениями. В сценарии залогового удержания подразумевается закрытие активов A в цепи X при наличии условий блокировки в зависимости от активности в цепи Y. Из этого следует, что все проходит в два этапа: (1)Блокировка залога на время использования активов; (2)Возврат залога пользователю через некоторое время при условии возвращения активов обратно их владельцу. Созданный алгоритм реализуется на платформе Ethereum [3]. Код для smart-контрактов пишется на языке Solidity. Тестирование алгоритма проводится на языке Python.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

---

[1] В. Бутерин, «Chain Interoperability». R3 Research, Сентябрь 9, 2016.

[2] Морис Херлихи, «Atomic Cross-Chain Swaps», PODC'18, Июль 23-27, 2018, Эгхем, Великобритания

[3] А. Левис, «A Gentle Introduction to Ethereum», Октябрь 2, 2016.

Научные руководители – к.ф.-м.н. Н.Н.Токарева, П.А.Сазонова

УДК 519.7

Связь кватернарных и булевых бент-функций

А. С. Шапоренко
Институт математики им. С. Л.Соболева СО РАН, г. Новосибирск
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research, г. Новосибирск

В [1] были определены q-арные ($g : \mathbb{Z}_q^n \to \mathbb{Z}_q$) бент-функции для $q > 1$. Изучение таких функций было обусловлено желанием авторов обобщить результаты работы [2] о применении булевых бент-функций в системах CDMA (Code Division Multiple Access). В настоящей работе исследуется связь кватернарных и булевых бент-функций.

Утверждение 1. *Пусть $g(x+2y) = a(x,y)+2b(x,y)$, где $x, y \in \mathbb{Z}_2$ и a и b — булевы функции от 2 переменных, является кватернарной бент-функцией, тогда b и $a \oplus b$ — бент-функции.*

Теорема 1. *Пусть $g(x+2y) = a(x,y)+2b(x,y)$, где $x, y \in \mathbb{Z}_2^n$ и a и b — булевы функции от $2n$ переменных, является кватернарной бент-функцией, тогда b и $a \oplus b$ — нелинейные функции при любом n.*

Утверждение 2. *Кватернарная функция от n перемнных $g(x + 2y) = a(x,y) + 2b(x,y)$, которая не является бент-функцией, тогда как b и $a \oplus b$ — булевы бент-функции от $2n$ перемнных, существует для любого n.*

Утверждение 3. *Кватернарная функция от n перемнных $g(x + 2y) = a(x,y) + 2b(x,y)$, которая является бент-функцией, тогда как b и $a \oplus b$ не являются бент-функциями от $2n$ переменных, существует для любого $n > 1$.*

Работа выполнена при подддержке лаборатории криптографии JetBrains Research и РФФИ (18-07–01394).

[1] Kumar P. V., Scholtz R. A., Welch L. R. Generalized bent functions and their properties. J. Combin. Theory Ser. A 40. 1985. P. 90 — 107.

[2] Olsen G. D., Scholtz R. A., Welch L. R. Bent-function sequences. IEEE Trans. Inform. Theory(1982) P. 858 — 864.

Научные руководители — к. ф.-м. н, доц. Н. Н. Токарева, А. В. Куценко