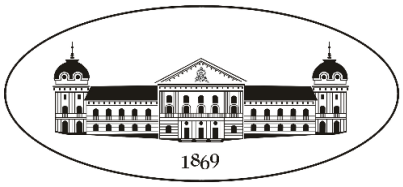
Конструиране на булеви функции и цифрови последователности за криптологията и комуникациите



Мирослав Маринов Димитров

Научен ръководител: проф. дн Цонка Байчева

Институт по Математика и Информатика
Българска Академия на Науките
Секция "Математически Основи на Информатиката"

автореферат на дисертация
за получаване на образователната и научна степен доктор
професионално направление 4.6
информатика и компютърни науки

София 2023

Благодарности

Искам да изразя дълбоката си благодарност към моя научен ръководител проф. дн Цонка Байчева за нейната безценна обратна връзка, търпеливост и градивна критичност. Искам да благодаря на д-р Николай Николов за многобройните интересни беседи по темите от дисертацията, както и за това, че насочи моето внимание към проблемите свързани с апериодичните свойства на двоичните редици. Задължен съм на проф. Bernhard Esslinger от университета в Зиген, за неговата редакторска помощ и подкрепа, както и за безвъзмездно изпратените от него хардуерни части, които значително подпомогнаха разрешаването на няколко проблема свързани с ротационните редици. Благодарен съм на д-р Георги Иванов за неговата мотивация и принос да стана докторант, запознавайки ме с проф. дн Цонка Байчева. Специални благодарности към д-р Виолета Андреева за нейната безрезервна подкрепа. Накрая, но не на последно място, искам да благодаря на моето семейство, и най-вече на моите родители, които с примера и уроците си запалиха в мен любовта към науката. Благодаря на моята съпруга Гери, както и на дъщеря ни Марая, за нейната емоционална подкрепа и вдъхновение.

0.1 Научни приноси

Главните научни приноси могат да бъдат обобщени така:

1. Подробно е анализирана богата колекция от популярни векторни булеви функции.

2. Демонстрирано е, че една голяма част от векторните булеви функции построени посредством функции от теория на хаоса са уязвими на линеен криптоанализ. Предложен е ефикасен алгоритъм, който достига до значително по-добри характеристики от тези, които авторите на гореспоменатите векторни булеви функции използват за сравнение.

3. Въведени са нови помощни структури и дефиниции, като куплунги, декомпозиция по координати и CELAT таблица, които позволяват формулирането на проблема по нелинейната оптимизация на булеви функции като фамилия от SAT проблеми.

4. Открити са 8×8 биективни векторни булеви функции съставени от координати с максимална нелинейност 116.

5. Предложена е стратегия за спектрален анализ на векторни булеви функции за откриване на аномалии.

6. Предложен е ефикасен евристичен алгоритъм за конструиране на двоични редици с малки пикове на страничните апериодични автокорелационни листове (PSL). Бяха намерени рекордни PSL стойности за дължини на редици между 106 и 300.

7. Предложено е подобрение на гореспоменатия алгоритъм за свеждане на времевата сложност и изискуемата памет до линейни стойности. Направените модификации позволяват достигането на рекордни PSL стойности за по-малко от секунда, дори и при редици с по-големи дължини.

8. Направен е подробен анализ и сравнение на ефикасността на алгоритмите при различни входни параметри. Това позволи създаването на нов алгоритъм, който успешно достигна до оптималните PSL стойности за известните от литературата дължини, които са открити след пълно обхождане. За целта беше използвана среден клас компютърна станция, като необходимото време за достигането на оптималните стойности за повечето от дължините е по-малко от секунда.

9. Предложен е алгоритъм за видео ускорители, който е насочен към намиране на оптимални PSL стойности измежду дадена двоична редица B и всички нейни възможни ротации. Намерени са всички PSL оптимални m-редици, с или без ротации, за дължини $2^n - 1$, за $18 \leq n \leq 20$. Намерени са и всички PSL оптимални редици на Legendre, с или без ротации, до дължини 432100. Резултатите водят до предположението, че всички PSL оптимални редици на Legender, с или без ротации, и за дължини N по-големи от 235723, притежават PSL стойност строго по-голяма от $\sqrt{N}$.

10. Предложени са няколко подобрения на водещите алгоритми за конструиране на изкривено-симетрични (skew-symmetric) двоични редици с ниски качествени стойности (MF), които редуцират сложността по памет от n^2 до n, без това да повлиява на времевата им сложност. Предложен е и алгоритъм, който оптимизира случайно генерирани изкривено-симетрични двоични редици с дължини до $10^5 + 1$ и MF стойности строго по-големи от 5. Това противоречи на предположението на Bernasconi, че няма да бъде създадена стохастична процедура, която да намира двоични редици с дължини над 200 и MF стойности над 5.

11. Предложени са нови класове двоични редици с четни дължини и алтерниращи автокорелационни странични листи равни на 1 по абсолютна стойност. MF стойностите на предложения клас редици са близки до MF стойностите на изкривено-симетричните редици открити от Golay.

12. Предложени са подкласове от редици, базирани на проблема с разбиването на числа. Открити са двоични редици с рекордни MF стойности за много от дължините до 225 и за всички дължини над 225. Като допълнителна демонстрация за ефикасността на предложения алгоритъм, той беше стартиран за оптимизация на двоични редици с дължини 573 и 1009. За тях е установено, че за един от водещите алгоритми са нужни съответно 32 и 46774481153

години за достигането на MF стойност по-голяма от 6.34. Предложеният в дисертацията алгоритъм достига тези MF стойности за тези дължини в рамките на няколко часа.

13. Предложена е нова стратегия за спектрален анализ на векторни булеви функции посредством апериодични автокорелационни функции.

0.2 Публикации по дисертацията

1. Dimitrov, Miroslav M. "On the design of chaos-based S-boxes."IEEE Access 8 (2020): 117173-117181, IF:3.367, Q2.

2. Dimitrov, Miroslav, Tsonka Baitcheva, and Nikolay Nikolov. "Efficient generation of low autocorrelation binary sequences."IEEE Signal Processing Letters 27 (2020): 341-345, IF:3.109, Q2.

3. Dimitrov, Miroslav, Tsonka Baitcheva, and Nikolay Nikolov. "On the generation of long binary sequences with record-breaking PSL values."IEEE Signal Processing Letters 27 (2020): 1904-1908, IF:3.109, Q2.

4. Dimitrov, Miroslav. "On the aperiodic autocorrelations of rotated binary sequences."IEEE Communications Letters 25.5 (2020): 1427-1430, IF:3.436, Q2.

5. Dimitrov, Miroslav, Tsonka Baicheva, and Nikolay Nikolov. "Hybrid Constructions of Binary Sequences With Low Autocorrelation Sideobes."IEEE Access 9 (2021): 112400-112410, IF:3.476, Q2.

6. Dimitrov, Miroslav M. "A Framework for Fine-Grained Nonlinearity Optimization of Boolean and Vectorial Boolean Functions."IEEE Access 9 (2021): 124910-124920, IF:3.476, Q2.

7. Iliev, M., Nikolov, N., Dimitrov, M. and Bedzhev, B. "Genetic algorithm for synthesis of binary signals with optimal autocorrelation."2020 International Conference on Information Technologies (InfoTech). IEEE, 2020.

8. Dimitrov, Miroslav. "On the Skew-Symmetric Binary Sequences and the Merit Factor Problem."arXiv preprint arXiv:2106.03377 (2021).

9. Dimitrov, Miroslav. "New Classes of Binary Sequences with High Merit Factor."arXiv preprint arXiv:2206.12070 (2022).

Глава 1

Увод

Булевите функции, векторните булеви функции (С-кутии) [1] и цифровите редици са широко използвани в различни области като телекомуникациите, радарните технологии, навигацията, криптографията, измервателните науки, биологията и в различни сфери на индустрията.

С-кутиите са един от най-важните криптографски примитиви в модерните блок шифри. Една уязвима С-кутия може да бъде атакувана с разнообразни методи като линеен криптоанализ [16], диференциален криптоанализ [17] , бумеранг атаки [140], алгебрични атаки [33] и други [54]. Една от най-важните характеристики за дадена С-кутия е нейната нелинейност. Висока нелинейност може да бъде достигната с метода на обратимите елементи над крайни полета [106], но съществуват опасения, че така генерираните С-кутии могат да бъдат уязвими към алгебрични атаки. Поради тази причина, в литературата са предложени различни стратегии за оптимизиране на случайно генерирани С-кутии до С-кутии с високи нелинейности [31], [79], [100], [101] и [138].

Поради критичността на С-кутиите, от гледна точка на сигурността на алгоритъма в който те са използвани, те трябва да бъдат придружени от подробна техническа документация описваща процеса и метода по тяхното конструиране. За съжаление, в редица случаи такава документация липсва, или по-лошо, дадената С-кутия е придружена с невалидна такава. Например, първоначалните С-кутии използвани в алгоритъма DES [50] са модифицирани от националната агенция по сигурност (NSA) на САЩ. Причините зад тази модификация не са били известни. По-късно обаче, в [32], D. Coppersmith ги оповестява. Оказва се, че агенцията е била наясно за съществуването на диференциалните атаки 20 години преди академичния свят.

[1] S-boxes, С от субституционни

Липсата на документация описващата създаването на дадена C-кутия може да бъде свързано и с наличието на скрита структура в нея, която дава на създателите значително предимство при хардуерната имплементация. Например, както е показано в [20], C-кутиите използвани от хеш функцията Streebog и 128 битовия блок шифър Kuznyechik, стандартизирани от Руската Федерация, са създадени с такава скрита структура.

От практическа гледна точка, липсата на техническа документация за това как дадена C-кутия е била конструирана може да бъде свързано с наличие на скрита структура, като тази предложена в [121]. Въпреки че наличието на такава структура лесно може да бъде установено, както е показано в [144], намирането на друг вид техники за маскиране на структури не трябва да се подценява. Нещо повече, такава структура може да се породи и неумишлено, което още повече подчертава важността на проблемите свързани с реверсивното инженерство на C-кутиите.

Намирането на двоични редици с колективно малки апериодични автокорелационни характеристики по предварително зададен критерий е известен и добре изследван проблем. Примери за такива критерии са нивото на максималния страничен лист (Peak Sidelobe Level - PSL) и качествения фактор (Merit Factor – MF), предложен от Golay през 1972 г. [55].

От практическа гледна точка, най-желаното свойство на дадена двоична редица е наличието на нисък PSL. Някои добре известни конструкции на такива редици включват кодовете на Barker [8], редиците на Rudin-Shapiro [122][129], m-редици [62], кодове на Gold [61], кодове на Kasami [78], редици на Weil [123], редици на Legendre [117]. Въпреки наличието на многобройни конструкции, нито една от тях не гарантира достигането на минимална (оптимална) PSL стойност за дадена фиксирана дължина. Оказва се, че единственият възможен подход за гарантиране на оптималност е чрез пълно обхождане. За съжаление, големината на обхожданото пространство нараства експоненциално с големината на редицата. Поради тази причина, не е изненадващо че PSL-оптималните стойности за двоични редици с дължини по-големи от 84 са все още неизвестни.

Golay отделя на изследването на MF-проблема близо 20 години (повече информация за това може да бъде намерена в [74]). След този период, биват публикувани разнообразни подходи и стратегии за конструкции на двоични редици с висока MF стойност, които включват евристични и алгебрични методи. В [60], MF-проблемът бива описан от Golay като "предизвикателен и очарователен".

Задачата за минимизирането на MF е известна и като LABS проблем. Той е тясно обвързан с проблеми в теоретична физика и химията. Например, LABS проблемът е свързан с квантовия модел на магнетизма. В [13] Bernasconi изказва предположението, че "търсенето чрез стохастични алгоритми няма да доведе до получаването на дълги двоични редици с MF стойност по-голяма от 5". Под "дълги двоични редици"Bernasconi подразбира двоични редици с дължини по-големи от 200. В [39], този проблем бива описан като "един от най-трудните оптимизационни проблеми".

Настоящият дисертационен труд разглежда няколко различни стратегии за конструиране и анализ на булеви функции, C-кутии и цифрови редици. В Глава 2 (раздели 2.1 и 2.2) са включени дефинициите и означенията, които се използват в последващите секции и глави. В Раздел 2.3 подробно са анализирани някои популярни C-кутии. Те могат да бъдат разделени на четири основни типа конструкции, както е показано в Раздел 2.4. C-кутии генерирани чрез помощта на функции от теорията на хаоса (ХФ) са детайлно анализирани, с цел измерване тяхната устойчивост към линейния криптоанализ. Голяма част от публикуваните трудове, свързани с ХФ, разглеждат само средната стойност на нелинейността на координатите на дадената C-кутия, игнорирайки останалите компоненти. Използването на такива C-кутии в конкретна криптосистема трябва да се извършва с повишено внимание. Нещо повече, оказва се че в контекста на нелинейната оптимизация, ползата от използването на ХФ е пренебрежима. Предложените два евристични алгоритъма, които стартират от напълно случайно генерирана C-кутия, достигат рекордни средни стойности на нелинейност, изчислена само върху координатите на C-кутията. В Раздел 2.5, проблемът свързан с нелинейната оптимизация на C-кутиите е формулиран като проблем за удовлетворимост, който може да се реши чрез използването на SAT solvers. Това се постига чрез въвеждането на нови помощни понятия и характеристики като куплунг, декомпозиция по координати, степен на спускаемост, разширена апроксимираща линейна таблица по координати (CELAT) и др. Оказва се, че балансирана C-кутия с размери 8x8 може да се построи от булеви функции на осем променливи като всяка от тях е с нелинейност 116. SAT апаратът може да бъде полезен в ситуации, когато създателят на дадената C-кутия иска да увеличи (или умишлено намали) нелинейността ѝ, прилагайки възможно най-малко промени по нейната структура. Например, нелинейностите на C-кутиите Skipjack, предложена от националната агенция по сигурност на САЩ и Kuznyechik, предложена от стандартизиращата агенция на Руската Федерация,

могат да бъдат оптимизирани до по-висока нелинейност чрез модифицирането на съответно 4 и 12 бита (от общо 2048).

В Глава 3 се разглежда стратегия за намиране на аномалии в структурите на C-кутиите чрез използването на спектрален анализ, разширявайки тази разгледана в [112].

Глава 4 е изцяло ориентирана към PSL проблема. В Раздел 4.1 е предложен ефикасен алгоритъм за конструиране на двоични редици, с помощта на който бяха достигнати рекордни PSL стойности на двоични редици с дължини между 106 и 300. В Раздел 4.2 е предложен друг алгоритъм, чиято линейна сложност позволява намирането на рекордни PSL стойности, приложим и за редици с по-голяма дължина. За голяма част от случаите, достигането до тези рекордни стойности отнема по-малко от секунда. В Раздел 4.3 се разглеждат различни параметри заложени в оценящата функция на алгоритмите от предните раздели. Изводите от анализа позволяват да се конструират алгоритми, които за сравнително кратко време могат да достигнат до оптимални PSL стойности. Описани са и хибридни алгоритми, които използват някои добре познати алгебрични конструкции. Накрая, в Раздел 4.3.3, се разглежда един известен изчислителен проблем за намирането на най-ниската PSL стойност измежду дадена редица B и всички възможни ротации на B. Използвайки няколко полезни математически свойства, в раздела се демонстрира как този проблем може да се атакува от перфектно балансиран паралелен алгоритъм. Чрез пълно обхождане бяха открити оптималните PSL стойности на m-редиците с дължини $2^n - 1$, за $18 \leq n \leq 20$, както и пълен списък на оптималните PSL стойности на редиците на Legendre, с и без ротация, с дължини до 432100. Извършените изчисления и резултатите от тях, водят до предположението, че оптималната PSL стойност на редиците на Legendre, с или без ротация, за дължини N по-големи от 235723, са строго по-малки от $\sqrt{N}$.

Глава 5 разглежда MF проблема. В Раздел 5.1 се разглеждат няколко полезни математически свойства, които описват връзката между дадена изкривено-симетрична двоична редица B и получената от нея редица посредством промяната на точно 2 елемента. Изведените свойства, позволяват изискуемата памет за съществуващите алгоритми да бъде редуцирана от n^2 до n. Предложеният алгоритъм успешно достига MF стойности над 5 за двоични редици с дължини до $10^5 + 1$.

Заради допълнителните свойства на изкривено-симетричните двоични редици, голяма част от трудовете, търсещи двоични редици с рекордни MF стойности, се фокусират върху тях. Трябва да се отбележи обаче, че те са дефинирани само за редици с нечетна дължина. Поради тази причина двоичните редици с четна дъл-

жина са често пренебрегвани. В Раздел 5.1.2 се предлага нова фамилия от двоични редици с четни дължини и алтернириращи абсолютни стойности на страничните листове равни на 1. Оказва се, че предложеният клас е тясно свързан с класа на изкривено-симетричните редици. Въвеждат се и няколко помощни подкласа редици, породени от задачата за разбиване на числа и потенциали измерени чрез помощни троични редици. Открити са двоични редици от предложения клас с рекордни MF стойности за някои дължини по-малки от 225, и за всички останали дължини над 225. Открити са и двоични редици, с четни и нечетни дължини, по-малки от 2^8 и MF стойност по-голяма от 8, както и четни и нечетни дължини, по-малки от 2^9 и MF стойност по-голяма от 7. Ефективността на предложения алгоритъм може да се демонстрира и при стартирането му върху две значително трудни за обхождане пространства включващи двоичните редици с дължини 573 и 1009. В [23], Фигура 7, е изчислено необходимото приблизително време нужно на алгоритъма lssOrel_8 за достигане на двоична последователност с MF стойност по-голяма от 6.34. Оказва се, че при редици с дължина 573 и 1009 необходимото време е съответно 32 и 46774481153 години. С предложеният алгоритъм, ние достигаме тези стойности за по-малко от няколко часа. Последният Раздел 5.2 предлага техника за спектрален анализ на C-кутии посредством визуализация на страничните листи.

Глава 2

Векторни булеви функции в криптографията

2.1 Булеви функции

Дефиниция 2.1.1 (Булеви функции & Таблици за истинност). Нека вземем множеството $B = \{0,1\}$. Булева функция $f(x)$ с n променливи $x_1,...,x_n$ е проекция $f : B^n \mapsto B$ от n двоични входа $x = (x_1,x_2,\cdots,x_n) \in B^n$ към един двоичен изход $y = f(x) \in B$. Двоичната таблица за истинност (BTT) на булева функция с n променливи $f(x)$ се състои от вектора съставен от всички последователни изхода на булевата функция:

$$[f(x)] = [f(0,0,\cdots,0), f(0,0,\cdots,1), \cdots, f(1,1,\cdots,1)]$$

Полярната таблица за истинност (PTT) на булева функция с n променливи $f(x)$ се получава директно от BTT. Дефинираме PTT със следната връзка: $[\hat{f}(x)] = [1-2f(x)]$.

Дефиниция 2.1.2 (Алгебрична нормална форма). Алгебричната нормална форма на булева функция с n променливи $f(x)$, или ANF_f, се дефинира със следното уравнение: $ANF_f = a_0 \oplus a_1x_1 \oplus a_2x_2 \oplus \cdots \oplus a_nx_n \oplus a_{1,2}x_1x_2 \oplus \cdots \oplus a_{1,2,\cdots,n}x_1x_2\cdots x_n$, където коефициентите a принадлежат на B.

Дефиниция 2.1.3 (Алгебрична степен). Алгебричната степен на една булева функция $f(x)$, или $deg(f)$, е равна на броя на променливите в най-дългия моном от ANF_f.

Дефиниция 2.1.4 (Разстояние по Hamming). Разстояние по Hamming между две булеви функции с n променливи $f(x)$ и $g(x)$, или $d_H(f,g)$, е равно на броя на различните елементи в техните съответни таблици за истинност.

Дефиниция 2.1.5 (Линейна булева функция). Всяка булева функция с n променливи от вида:

$$l_w(x) =< w,x >= w_1x_1 \oplus w_2x_2 \oplus \cdots \oplus w_nx_n,$$

където $w,x \in B^n$, се нарича линейна функция.

Дефиниция 2.1.6 (Афинна булева функция). Всяка булева функция с n променливи от вида:

$$l_w(x) =< w,x >= w_0 \oplus w_1x_1 \oplus w_2x_2 \oplus \cdots \oplus w_nx_n,$$

където $w_0 \in B$ и $w,x \in B^n$, се нарича афинна булева функция.

Дефиниция 2.1.7 (Трансформация на Walsh-Hadamard). За една булева функция с n променливи $f(x)$, представена чрез PTT, $[\hat{f}(x)]$, трансформацията на Walsh-Hadamard, или WHT, $\hat{F}_f : B^n \to Z$, е дефинирана така:

$$\hat{F}_f(w) = \sum_{x \in B^n} \hat{f}(x)(-1)^{<w,x>}$$

Дефиниция 2.1.8 (Абсолютен индикатор). За една булева функция с n променливи $f(x)$, бележим абсолютния индикатор на f като Δ_f. За всички $u \in F_2^n$, освен нулевия вектор, имаме

$$\Delta_f(u) = \sum_x (-1)^{f(x)+f(x+u)}$$

Абсолютният индикатор на f е равен на:

$$\Delta_f = \max_u \mid \Delta_f(u) \mid \tag{2.1}$$

2.2 Векторни булеви функции (C-кутии)

Дефиниция 2.2.1 (Векторна булева функция – Субституционна таблица – C-кутия – S-box). Проекция от n на брой двоични входа към m на брой двоични изхода $S : B^n \Leftrightarrow B^m$, която определя някое $y = (y_1, y_2, \cdots, y_m) \in B^m$ като $S(x) = y$ за всяко $x = (x_1, x_2, \cdots, x_n) \in B^n$, се нарична (n,m) субституционна таблица (C-кутия; S-box) и се бележи с $S(n,m)$.

Дефиниция 2.2.2 (Биективна (Балансирана) С-кутия). С-кутия $S(n,m)$ се нарича биективна, ако всеки вход $x \in B^n$ е свързан с различен изход $y = S(x) \in B^m$ и всички възможни 2^m на брой изхода са налични.

Дефиниция 2.2.3 (Референтна таблица – S-box Look-up Table – LUT). Референтната таблица LUT на една С-кутия $S(n,m)$ е (2^n x m) двоична матрица S, чиито редове са съставени от всички възможни изходи на $S(n,m)$, съответстващи на всички възможни 2^n на брой входа, наредени лексикографски.

Дефиниция 2.2.4 (Координати). Дефинираме всяка колона от $S(n,m)$ LUT като координата на $S(n,m)$.

Дефиниция 2.2.5 (Полярна референтна таблица – Polarity Look-up Table – PLUT). Полярната референтна таблица PLUT на С-кутия $S(n,m)$, или S_{PLUT}, е $(2^n,m)$ матрица с елементи от $\{-1,1\}$, където всеки елемент от ред j и колона k, отбелязан с $S_{PLUT}[j][k]$, за $j = 1,2,\cdots,2^n$ и $k = 1,2,\cdots,m$, се получава от $S_{LUT}[j][k]$ като

$$S_{PLUT}[j][k] = (-1)^{S_{LUT}[j][k]} = 1 - 2S_{LUT}[j][k]$$

където $\hat{f}_i(\alpha) = (-1)^{f_i(\alpha)} = 1 - 2f_i(\alpha)$.

Дефиниция 2.2.6 (Разширена WHT спектрална матрица – EWHTSM). Разширената Walsh-Hadamard спектрална матрица (EWHTSM) на С-кутия $S(n,m)$ е $(2^n,2^m)$ матрица $\hat{F}_{ExtS}$, с колони образувани от трансформацията по Walsh-Hadamard $[\hat{F}_{g_v}(w)]$ на булевите функции $g_v(x) = v_1 f_1(x) \oplus v_2 f_2(x) \oplus \cdots \oplus v_m f_m(x)$, където w и v са подредени лексикографски съответно в B^n и B^m.

$$\hat{F}_{ExtS} = \begin{bmatrix} \hat{F}_{g_0}(0,0,\cdots,0) & \cdots & \hat{F}_{g_{2^m-1}}(0,0,\cdots,0) \\ \hat{F}_{g_0}(0,0,\cdots,1) & \cdots & \hat{F}_{g_{2^m-1}}(0,0,\cdots,1) \\ \vdots & \ddots & \vdots \\ \hat{F}_{g_0}(1,1,\cdots,0) & \cdots & \hat{F}_{g_{2^m-1}}(1,1,\cdots,0) \\ \hat{F}_{g_0}(1,1,\cdots,1) & \cdots & \hat{F}_{g_{2^m-1}}(1,1,\cdots,1) \end{bmatrix} \tag{2.2}$$

Разширената спектрална матрица ни дава една количествена мярка, като например разстоянието по Hamming, с която да измерим разстоянията между всички линейни комбинации от координати на дадена С-кутия и всички възможни линейни функции от същата размерност.

Дефиниция 2.2.7 (Линейна апроксимираща таблица – LAT). Линейната апроксимираща таблица на С-кутия $S(n,m)$, или LAT_S, или S_{LAT}, е таблица с 2^n реда и 2^m

колони, определена така:

$$S_{LAT}[X][Y] = LAT_S[X][Y] = 2^{n-1} - d_H(X,Y), \tag{2.3}$$

където Y е поредната линейна комбинация от координати на С-кутията, а X е поредната линейна функция над n.

Дефиниция 2.2.8 (Нелинейност). Нелинейността на дадена С-кутия $S(n,m)$, или S_{NL}, е дефинирана така:

$$S_{NL} = 2^{n-1} - \max(\{\mid w_i \mid\}), \tag{2.4}$$

където $\{\mid w_i \mid\}$ е множеството от всички елементи на LAT по абсолютна стойност, без най-горния ляв елемент.

Дефиниция 2.2.9 (Средна координатна нелинейна стойност (ACNV) на С-кутия). Средната координатна нелинейна стойност, или S_{ACNV}, на дадена С-кутия S, е средната стойност от нелинейностите на координатите на S.

Дефиниция 2.2.10 (Десетична референтна таблица – DLUT). Всяка С-кутия е еднозначно определена от нейната LUT. Трансформирайки всеки ред от LUT към десетично число еднозначно се определя същата С-кутия и чрез десетична референтна таблица (DLUT).

Дефиниция 2.2.11 (XOR Таблица). XOR таблицата на дадена С-кутия $S(n,m)$ е $(2^n \times 2^m)$ двоична матрица S_{XORT}, чиито колони са формирани от всички линейни комбинации от колони от S_{LUT} подредени лексикографски.

Дефиниция 2.2.12 (Минимална алгебрична стойност). Минималната алгебрична стойност на дадена С-кутия $S(n,m)$ е минималната алгебрична степен измежду всички компонентни функции на S.

$$\begin{aligned} S_{DEG} &= min_{(v \in B^m)} deg(g_v) = \\ &= min_{((v_1,v_2,\cdots,v_m) \in B^m)} deg(v_1 f_1(x) \oplus v_2 f_2(x) \oplus \cdots \oplus v_m f_m(x)), \end{aligned} \tag{2.5}$$

където $f_1, f_2, \cdots, f_m$ са координатните булеви функции на $S(n,m)$.

Дефиниция 2.2.13 (Абсолютен индикатор). Абсолютният индикатор на дадена С-кутия S, или S_{AC}, е равен на максималния абсолютен индикатор измежду всички абсолютни индикатори на компонентни функции на S.

Дефиниция 2.2.14 (Диференциално еднообразие – Differential Uniformity). Диференциалното еднообразие, или δ на дадена С-кутия $S(n,m)$, бележим с S_δ, и дефинираме така:

$$S_\delta = \max_{\alpha \in B^n \setminus \{0\}} \max_{\beta \in B^m} \left| \{ x \in B^n \mid S(x) \oplus S(x \oplus \alpha) = \beta) \} \right|$$

2.3 Криптографски свойства на някои популярни С-кутии

Криптографските свойства на векторните булеви функции са подробно разгледани в научната литература. Известни са желаните параметри и характеристики, които да гарантират надеждност на дадената С-кутия при атаки тип линеен криптоанализ [96][16], диференциален криптоанализ [17], бумеранг-атака [140] или интерполационна атака [73]. С-кутиите са важен криптографски примитив широко използван в модерни криптографски алгоритми като AES [38], Whirlpool [10], Camellia [6] и др. За дадена С-кутия S, целта на дизайнера е да постигне високи стойности на S_{NL} и S_{DEG}, както и ниски такива на S_δ и S_{AC}.

С-кутиите, създадени с методът на обратимите елементи над крайни полета [107], като С-кутията Rijndael използвана в AES [38], имат най-добрите криптографски характеристики измежду всички известни 8×8 кутии. Съществуват обаче опасения, че такива кутии могат да бъдат уязвими към алгебрични атаки [33]. Поради тази причина, в редица приложения се използват случайно или евристично генерирани С-кутии. В този раздел е събрана колекция от добре известни С-кутии, за да бъдат подробно анализирани. Оказа се, че само 11 от 47-те анализирани кутии са AES-подобни. За по-детайлна картина е представена и LAT спектралната картина на всяка една от тях. Разпределението на S_{LAT} коефициентите може да бъде полезно от гледна точка на реверсивното инженерство.

2.4 Стратегии за конструиране на С-кутии

Разнообразните методи за конструиране на С-кутии могат да бъдат разпределени в четири основни категории. Първата категория $\boldsymbol{T_1}$ за намиране на С-кутии с добри криптографски свойства е изцяло псевдо-случайна. Най-високата постигната нелинейност на $(8,8)$ кутия генерирана чрез този метод е 100 [103].

Втората категория $\boldsymbol{T_2}$ включва методи с по-директен подход, като алгебрични конструкции, като тази на крайните елементи над крайни полета, автомати [15],

квази-циклични кодове [24][18], APA методи [37] или Feistel и Misty конструкции [28].

Третата категория $\boldsymbol{T_3}$ включва методи, които прилагат евристични алгоритми за оптимизирането на случайно генерирани C-кутии. Пример за такива са hill climbing [100], simulated annealing [31], генетични алгоритми [101], специални генетични алгоритми [138], имунни алгоритми [72], и други [135][114].

Четвъртата категория $\boldsymbol{T_4}$ се състои от хибридни методи – методи, които започват оптимизационният процес от кутия генерирана с метод от категория T_2, след което прилагат алгоритъм от категория T_3. Примери за такива методи са [79], [30], [70], [94], [40], [71] и [4].

Подробно са разгледани и методите за конструиране на C-кутии посредством функции от теория на хаоса. Те могат да принадлежат на категориите T_2, T_3 или T_4. В [45], голяма част от тези кутии биват анализирани, за да се измери тяхната реална устойчивост към линейния криптоанализ. Повечето от авторите, използващи теория на хаоса, считат за най-важна характеристиката ACNV, игнорирайки нелинейността на компонентните функции на кутията. Интегрирането на такива кутии в критични приложения трябва да става с повишено внимание. Нещо повече, в някои от случаите се оказва, че ползата от използването на хаотични функции като метод за конструиране на C-кутии е пренебрежимо малка. Използвайки два евристични метода и започвайки от случайно генерирани C-кутии, демонстрираме как бързо и ефикасно може да се достигат C-кутии с по-добри криптографски характеристики от тези базирани на хаотични функции (ХФ).

Методите използвани за генериране на C-кутии посредством ХФ са много и разнообразни (виж [45]). Според Дефиниция 2.2.8, ние търсим максималния елемент по абсолютна стойност v измежду всички елементи в LAT таблицата на C-кутията $S(n,n)$, за да определим нейната нелинейност, т.е. $S_{NL} = 2^{n-1} - v$. Както е показано в [96][16][68], ниска нелинейност се асоциира с висока вероятност за успешно прилагане на линеен криптоанализ. В [45] е показано, че ACNV не отразява реалната нелинейност на дадена C-кутия. В случите когато за дадено приложение ACNV стойността е по-значима от общата нелинейност, то предложената в тази работа евристична конструкция дава по-добри резултати от останалите публикувани в литературата.

Ако искаме да понижим нелинейността на дадена балансирана C-кутия $S(n,n)$, то стратегията свързана с колективното минимизиране на елементите от S_{LAT} по абсолютна стойност е логична. Нещо повече, елементите от всяка колона на S_{LAT} са обвързани с теоремата на Parceval [97]. Нека означим с C_i списъка от елементи

съставен от $S_{LAT}[i]$. Тъй като искаме да се фокусираме върху понижаване на нелинейността на координатите на S, предлагаме следната оценяща функция $E(S) = \sum_{p=0}^{n-1} \sum_{x \in C_{2^p}} |x|^M$, където с M бележим нейната мощност, а ограничението $x \in C_{2^p}$ ограничава нейния обхват.

Използвайки евристична функция с оценка $E(S)$, стартирайки от случайно генерирана C-кутия, почти винаги достигаме до C-кутия с ACNV стойност 114.0, най-високата публикувана в литературата. Нещо повече, прилагайки разширение на алгоритъма посредством идеите предложени в [14], може да се достигне и до стойност 114.5 (виж [45]). Алгоритъмът е реализиран посредством инструменти от SageMath [41].

2.5 Нелинейна оптимизация посредством SAT техники

В този раздел, показваме връзка между нелинейната оптимизация на булеви и векторни булеви функции с целочисленото програмиране.

Съществен недостатък в съвременните евристични техники за нелинейна оптимизация на C-кутии е тяхната агресивност към оптимизиращата се C-кутия. В повечето случаи е невъзможно да се намери връзка между началното и крайното състояние на оптимизирания обект. Оптимизационната техника предложена в [46] позволява нелинейната оптимизация да се извърши посредством минимални промени по оригиналната кутия. Това свойство може да бъде особено полезно, когато искаме да се фокусираме само в отслабващите нелинейността компоненти, без да влошаваме характеристиките на останалите. Ефективността на алгоритъма се демонстрира чрез примери за повишаването на нелинейността на C-кутиите Skipjack и Kuznyechik, с промяна единствено на съответно 4 и 12 бита (от общо 2048).

Максималната известна нелинейност на балансирана булева функция с 8 променливи е 116 [115]. Както е показано в [126], нелинейността е ограничена до 120, което показва че теоретичната максимална стойност на ACNV на (8,8) биективна C-кутия е по-малка или равна на 118.0. Ако се намери такава, то тя ще има координата с нелинейност 118, което ще даде отговор на въпроса дали съществува балансирана булева функция с нелинейност 118. Трябва да се отбележи, че академичните среди гледат скептично към съществуването на такава. Съществува ли обаче биективна C-кутия с ACNV стойност 116.0? Използването на SAT техника за нелинейна оптимизация на C-кутии даде положителен отговор на този въпрос.

За съжаление, със същия апарат не успяхме да достигнем до балансирана булева функция с нелинейност 118.

За реализацията на алгоритъма се използват някои полезни означения и дефиниции, като куплунги, декомпозиция по координати, степен на спускаемост и др. Нека обозначим и с $f(n)^i$ десетичната стойност на n, при промяна на i-тия бит в двоичното представяне на n.

Лема 2.5.1. Ако променим точно един бит в LUT таблицата на биективна C-кутия S, то ще нарушим нейното биективно свойство.

Лема 2.5.2. Най-малкият (не нулев) брой от битове, които трябва да променим в дадена биективна C-кутия, запазвайки нейното биективно свойство е 2.

Дефиниция 2.5.1 (Куплунги). Нека вземем биективна C-кутия $S(n,n)$ и нейната DLUT таблица:

$$S_{DLUT} = [d_0, d_1, \cdots, d_i, \cdots, d_{2^n-1}].$$

Дефинираме като куплунг всяко множество $\{d_s, f(d_s)^j\}$, а множеството от всички куплунги в S като $\{S\updownarrow\}$.

Лема 2.5.3. За биективна C-кутия $S(n,n)$:

$$|\{S\updownarrow\}| = n2^{n-1}.$$

Дефиниция 2.5.2. Дефинираме множеството $\{S\updownarrow^i\}$ като максималното подмножество от множеството на куплунгите на дадена биективна C-кутия S(n,n), което съдържа куплунги опериращи само върху колона i от S_{LUT}, т.е. куплунги от вида $\{d_x, f(d_x)^i\}$. Ще наричаме всяко едно такова максимално подмножество $\{S\updownarrow^i\}$ стълбово куплунгово множество опериращо върху колона i от S_{LUT}.

Следствие 2.5.1. За биективна C-кутия S(n,n), са изпълнени следните свойства:

- $\forall i \neq j, \{S\updownarrow^i\} \cap \{S\updownarrow^j\} = \varnothing$
- $\forall i, |\{S\updownarrow^i\}| = 2^{n-1}$
- $|\bigcup_{i=1}^{n} \{S\updownarrow^i\}| = n2^{n-1}$

Дефиниция 2.5.3 (Декомпозиция по координати). Нека S е (n,n) биективна C-кутия. Нека вземем произволен елемент с координати (x,y) от нейната линейна апроксимираща таблица S_{LAT}. Двоичното представяне на y е:

$$y_{(2)} = b_{n-1}2^{n-1} + b_{n-2}2^{n-2} + \cdots + b_1 2^1 + b_0 2^0$$

Декомпозицията по координати на елемент с координати (x,y), или $\Delta_{x,y}$, е множеството:

$$\Delta_{x,y} = \bigcup_{i=0,b_i\neq 0}^{n-1} \{b_i(n-i-1)\}.$$

Дефиниция 2.5.4 (Снимка на нелинейното ограничение). Дефинираме като снимка на нелинейното ограничение S_{NBS} на дадена биективна C-кутия $S(n,n)$ множеството от двойки съдържащи всички координати от елементи на S_{LAT}, които ограничават нелинейността S_{NL} на S, т.е

$$(x,y) \in S_{NBS} \Leftrightarrow |LAT_S[x][y]| = 2^{n-1} - S_{NL}.$$

Дефиниция 2.5.5 (Декомпозиция по координати на снимката). Дефинираме декомпозицията по координати на снимката на биективна C-кутия $S(n,n)$, или Δ_S, като множество от всички S_{NBS} декомпозиции по координати, т.е.:

$$\Delta_S = \bigcup_{(x,y)\in S_{NBS}} \Delta_{x,y}.$$

Дефиниция 2.5.6 (Степен на спускаемост – Λ_S). За дадена биективна C-кутия $S(n,n)$, дефинираме фамилия от множества Ψ_S, такава че:

$$E \in \Psi_S \Leftrightarrow \forall Q \in \Delta_S \ \exists q \in Q : \ q \in E.$$

Степента на спускаемост на S е равна на минималната мощност на множество в Ψ_S, т.е.:

$$\Lambda_S = \min_{\forall A\in\Psi_S} |A|.$$

Следствие 2.5.2 (Свойства на Λ_S). За дадена биективна C-кутия $S(n,n)$:

- $\Lambda_S \in \mathbb{N}$
- $\Lambda_S \in [1,n]$
- $\Lambda_S = 1 \Leftrightarrow |\bigcap_{s\in\Delta_S}| \geq 1$
- $\Lambda_S > 1 \Leftrightarrow \bigcap_{s\in\Delta_S} = \varnothing$

Дефиниция 2.5.7 (Спускаема координата). За дадена биективна C-кутия $S(n,n)$, казваме че координатата j е спускаема, ако са изпълнение следните свойства:

- $\Lambda_S = 1$

- $j \in \bigcap_{s \in \Delta_S}$

Дефиниция 2.5.8 (Куплунгува трансформация). За дадена биективна C-кутия $S(n,n)$ и куплунг c_i, бележим с S^{c_i} C-кутията получена след активирането на куплунг c_i върху S. Бележим тази куплунгува трансформация с оператора $\circ$, т.е.

$$S^{c_i} = S \circ c_i.$$

Когато имаме списък от куплунги $\{c_1, c_2, \cdots, c_i\}$, които искаме да активираме върху S в точно този ред, ще използваме следния израз:

$$S^{c_1,c_2,\cdots,c_i} = S \circ c_1 \circ c_2 \circ \cdots \circ c_i.$$

Лема 2.5.4. За дадена биективна C-кутия S и куплунг c е изпълнено следното свойство::

$$S = S \circ c \circ c.$$

Дефиниция 2.5.9 (Трансформираща куплунгова матрица – CTM). За дадена биективна C-кутия $S(n,n)$ и куплунг c_i, бележим с ${S^{c_i}}_{LAT}$ трансформираната LAT матрица на S след активацията на c_i. Трансформиращата куплунгова матрица на c_i върху S бележим така:

$$S^{c_i}_{CTM} = {S^{c_i}}_{LAT} - S_{LAT}.$$

Лема 2.5.5. За дадена биективна C-кутия $S(n,n)$, и за всеки два куплунга c_a и c_b, които принадлежат на едно и също куплунгово множество от стълбове $\{S\updownarrow^i\}$, имаме:

$$S \circ c_a \circ c_b = S \circ c_b \circ c_a.$$

Следствие 2.5.3. За дадена биективна C-кутия $S(n,n)$, за всички куплунги c_j, които принадлежат на едно и също куплунгово множество от стълбове $\{S\updownarrow^i\}$, са изпълнени следните свойства:

$$S^{c_a,c_b}_{LAT} = S^{c_b,c_a}_{LAT} = S_{LAT} + S^{c_a}_{CTM} + S^{c_b}_{CTM}$$

$$S^{c_1,c_2,\cdots,c_k}_{LAT} = S_{LAT} + \sum_{i=1}^{k} S^{c_i}_{CTM}$$

Лема 2.5.6 (CTM Стойности). Възможните стойности на елементите от CTM са -2, 0, или 2.

Следствие 2.5.4. За дадена биективна C-кутия $S(n,n)$, нека активираме куплунги $c_1, c_2, \cdots, c_k$, които принадлежат на едно и също куплунгово множество от стълбове $\{S \updownarrow^i\}$. Възможните стойности на изходната CTM са числа от интервала $[-2k, -2(k-1), \cdots, -2, 0, 2, \cdots, 2(k-1), 2k]$.

Дефиниция 2.5.10 (Координатно разширена LAT – CELAT). За дадена биективна C-кутия $S(n,n)$, и дадена координата i, дефинираме едномерната линейна апроксимираща таблица на S като:

$$S_{LAT_{1D}}[x] = S_{LAT}[x \ / \ 2^n][x \ \% \ 2^n].$$

Бележим всички куплунги от куплунговото множество от стълбове $\{S \updownarrow^i\}$ с $c_1, c_2, \cdots, c_{2^{n-1}}$. Тогава:

$$\begin{aligned} S^{c_1}_{CTM} &= S^{c_1}_{LAT} - S_{LAT} \\ S^{c_2}_{CTM} &= S^{c_2}_{LAT} - S_{LAT} \\ &\cdots \\ S^{c_{2^{n-1}}}_{CTM} &= S^{c_{2^{n-1}}}_{LAT} - S_{LAT}. \end{aligned} \tag{2.6}$$

По аналогичен начин можем да дефинираме и едномерната матрица CTM, т.е.:

$$\begin{aligned} S^{c_1}_{CTM_{1D}} &= S^{c_1}_{LAT_{1D}} - S_{LAT_{1D}} \\ S^{c_2}_{CTM_{1D}} &= S^{c_2}_{LAT_{1D}} - S_{LAT_{1D}} \\ &\cdots \\ S^{c_{2^{n-1}}}_{CTM_{1D}} &= S^{c_{2^{n-1}}}_{LAT_{1D}} - S_{LAT_{1D}}. \end{aligned} \tag{2.7}$$

Дефинираме i-та координатно разширена LAT S^i_{CELAT} по следния начин:

$$S^i_{CELAT} = \begin{bmatrix} S_{LAT_{1D}} \\ S^{c_1}_{CTM_{1D}} \\ S^{c_2}_{CTM_{1D}} \\ \cdots \\ S^{c_{2^{n-1}}}_{CTM_{1D}} \end{bmatrix}$$

S^i_{CELAT} има $2^{n-1}+1$ реда и 2^{2n} колони.

Дефиниция 2.5.11 (Двоично целочислено програмиране – SAT проблем). Задачата на двоичното целочислено програмиране е проблем от следния вид:

$$\begin{array}{lrcll} \text{рестрикции} & Ax & \leq & b & \\ & x & \geq & 0 & \text{двоично} \end{array}$$

където A е (m,n) матрица, а b и x са вектори със съответни дължини m и n. Векторът стълб x съдържа двоичните променливи, които трябва да бъдат оптимизирани. Казваме, че множеството S е множеството на допустимите решения, т.е.:

$$S := \left\{x \in B^n : Ax \leq b\right\}$$

В контекста на този проблем ние търсим само един елемент от множеството S, но не и оптималния такъв.

За всяка (n,n) C-кутия S, нека заменим 2^{n-1} с r и 2^{2n} с m. Нейната CELAT таблица за координата i е:

$$S^i_{CELAT} = \begin{bmatrix} S_{LAT_{1D}} \\ S^{c_1}_{CTM_{1D}} \\ S^{c_2}_{CTM_{1D}} \\ \cdots \\ S^{c_{2^n-1}}_{CTM_{1D}} \end{bmatrix} = \begin{bmatrix} l_1 & l_2 & \cdots & l_m \\ c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{r1} & c_{r2} & \cdots & c_{rm} \end{bmatrix}$$

Нека разгледаме случая, когато искаме да активираме куплунгите $P = p_1, p_2, \cdots, p_k$, които принадлежат на едно и също куплунгово множество от стълбове $\{S \updownarrow^i\}$. От следствие 2.5.3 следва:

$$S^{p_1,p_2,\cdots,p_k}_{LAT} = S_{LAT} + \sum_{i=1}^{k} S^{p_i}_{CTM}.$$

Бележим

$$S^{p_1,p_2,\cdots,p_k}_{LAT_{1D}} = [q_1, q_2, \cdots, q_m].$$

Тогава може да построим следната система от уравнения:

$$\begin{aligned} q_1 &= l_1 + c_{11}x_1 + c_{21}x_2 + \cdots + c_{r1}x_r \\ q_2 &= l_2 + c_{12}x_1 + c_{22}x_2 + \cdots + c_{r2}x_r \\ &\cdots \\ q_m &= l_m + c_{1m}x_1 + c_{2m}x_2 + \cdots + c_{rm}x_r. \end{aligned} \tag{2.8}$$

където $x = (x_1, x_2, \cdots, x_r) \in B^r$ и $x_t = 1$, ако $p_t \in P$. Имаме $S_{NL} = 2^{n-1} - \max_{j=1}^{m} abs(l_j)$. Ако координата i е спускаема, то тогава може да построим следната двоична целочислена програма:

$$\begin{array}{lrcll} \text{рестрикции} & < S^i_{CELAT}{}^T, x > & \leq & A & \\ \text{рестрикции} & < S^i_{CELAT}{}^T, x > & \geq & B & \\ & x & \geq & 0 & \text{двоично,} \end{array}$$

където A е вектор стълб с $2^{n-1}+1$ елемента, всеки равен на $2^{n-1} - S_{NL} - 2$, докато B е вектор стълб с $2^{n-1}+1$ елемента, всеки равен на $S_{NL} - 2^{n-1} + 2$. Бележим SAT проблемът свързан със спускането по координата i от уравнение 2.5 като $\Omega_{S,i}$. Това е проблем с общо 2^{n-1} двоични променливи и $2^n + 2$ рестрикции. Можем обаче да разделим този проблем на обединение от по-малки такива т.е.:

$$\Omega_{S,i} = \bigcup_{d=1}^{n-1} \Omega^d_{S,i}$$

където всеки подпроблем $\Omega^d_{S,i}$ е моделиран посредством следните ограничения:

$$\begin{array}{lrcll} \text{рестрикции} & < S^i_{CELAT}{}^T, x > & \leq & A & \\ \text{рестрикции} & < S^i_{CELAT}{}^T, x > & \geq & B & \\ \text{рестрикции} & \sum_{j=1}^{r} x_j & = & d & \\ & x & \geq & 0 & \text{двоично.} \end{array}$$

Разрешаването на който и да е от подпроблемите води и до решение на първоначалния проблем.

За подпроблемите $\Omega^d_{S,i}$ от $\Omega_{S,i}$ е изпълнено следното свойство::

$$\bigcap_{d=1}^{n-1} \Omega^d_{S,i} = \varnothing.$$

Също така, големината на множеството $\Omega^d_{S,i}$ за дадена биективна C-кутия $S(n,n)$ е $\binom{2^{n-1}}{d}$.

Теорема 2.5.1. За подпроблемът $\Omega^d_{S,i}$, всички ограничения с участието на някое l_j, за които следните неравенства са изпълнени:

$$\begin{aligned} l_j &\leq 2^{n-1} - S_{NL} - 2d - 2 \\ l_j &\geq S_{NL} - 2^{n-1} + 2d + 2 \end{aligned} \tag{2.9}$$

са винаги съвместими.

Дефиниция 2.5.12 (CELAT с радиус R). За дадена биективна C-кутия $S(n,n)$, и дадена координата i, имаме:

$$S^i_{CELAT} = \begin{bmatrix} l_1 & l_2 & \cdots & l_m \\ c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{r1} & c_{r2} & \cdots & c_{rm} \end{bmatrix}$$

Дефинираме с $S^{i,R}_{CELAT}$ матрицата построена от тези колони на S^i_{CELAT} с първи елемент ρ, за които са изпълнени следните неравенства::

$$\begin{aligned} \rho &> 2^{n-1} - S_{NL} - 2R - 2 \\ \rho &< S_{NL} - 2^{n-1} + 2R + 2. \end{aligned} \tag{2.10}$$

От тук следва, че всеки проблем $\Omega^d_{S,i}$ може да бъде редуциран до $S^{i,d}_{CELAT}$, вместо директното му (нередуцирано) разрешаване чрез S^i_{CELAT}.

Алгоритъмът е имплементиран посредством Python и Gurobi SAT Solver [65]. Бяха анализирани две известни C-кутии, Skipjack [131] и Kuznyechik [48], които успешно бяха оптимизирани до кутии с по-голяма нелинейност с промяна само на съответно 4 и 12 бита (от общо 2048).

2.5.1 Проблемът ACNV

Проблемът за нелинейната оптимизация на ACNV може да бъде разгледан като специален, сравнително олекотен, от гледна точка на изчислителни изисквания, случай. За входна C-кутия беше използвана кутията публикувана в [45] с ACNV стойност от 114.5. Тя се състои от 6 координати с нелинейност 114 и 2 координати с нелинейност 116. След сравнително кратко време, C-кутията беше оптимизирана

до кутия с ACNV от 116.0 и с обща нелинейност от 92. Всички опити да се достигне до C-кутия с по-висока ACNV стойност бяха неуспешни.

Глава 3

Реверсивно инженерство на C-кутии

3.1 Увод и мотивация

Причините за прикриването на използвания метод за създаването на дадена C-кутия са комплексни. Например, първоначалните C-кутии използвани в стандарта DES [50] биват модифицирани от националната агенция по сигурност на САЩ (NSA). Малко по-късно, D. Coppersmith разкрива каква е била мотивацията за тези промени [32]. Оказва се, че агенцията е разполагала с информация за съществуването на атаки тип диференциален анализ 20 години преди академичния свят. Coppersmith допълнително коментира защо са укрили тази информация казвайки:

> ... това беше защото [диференциалният криптоанализ] може да бъде много мощно средство, което може да се използва срещу редица криптосистеми, и имаше опасения, че предоставянето на тази информация на обществото ще се отрази негативно върху националната сигурност. [1]

Друга причина за прикриването на използвания метод за създаването на дадена C-кутия е наличие на скрита структура, която дава на създателя значимо предимство в нейното хардуерно реализиране. Например, както е показано в [20], C-кутиите използвани в хеш функцията Streebog и 128 битовия блок шифър Kuznyechik, стандартизирани от Руската Федерация, използват такива скрити структури. Потребител, който е наясно с непубликуваните спецификации, може

[1] ... that was because [differential cryptanalysis] can be a very powerful tool, used against many schemes, and there was concern that such information in the public domain could adversely affect national security.

хардуерно да реализира C-кутията по такъв начин, че обръщенията към нея да се извършват осем пъти по-бързо.

Още една причина за прикриването на използвания метод за създаването на дадена C-кутия може да бъде внедряване на зловредна структура, както е показано в [121]. Въпреки че конкретно тази структура може да бъде разпозната, както е показано в [144], мотивацията за откриване на метод за прикриване на зловредни структури от друг характер не трябва да бъде подценявана.

Съществуват различни техники и инструменти при прилагането на реверсивно инженерство на C-кутиите (виж [112][19][113]). В следващият раздел се разглеждат някои нови стратегии за спектрален анализ. Добър пример за прилагането на спектрален анализ е публикуван в [20].

3.2 Спектрален анализ на C-кутии

Можем да изолираме тези координати, представени като индекси на редове и колони, на елементите от LAT таблицата на дадена C-кутия $S(n,m)$, които са равни на дадена фиксирана стойност, или в по-общия случай, които принадлежат на даден интервал от стойности. Дефинираме всяко едно такова изолиране като спектрален канал. За удобство, бележим с $\S_S^E$ спектралния канал изолиран от C-кутия S, използвайки ограничителното множество E. Каналът може да се визуализира посредством $(2^n \times 2^m)$ матрична графика, като тези елементи които принадлежат на ограничителното множество са оцветени в червено, а останалите са безцветни.

По време на нашите експерименти, бяха случайно генерирани множество биективни C-кутии с размерност $(8,8)$, за да се проследи поведението на техните спектрални канали. Не бяха открити аномалии, видими симетрии или модели в някоя от тях. Всъщност, трудно беше да се различи визуално техния спектрален канал от случайно генериран канал с произволно разпръснати елементи.

В [125] е публикувана богата колекция от популярни C-кутии. В останалата част от раздела се предоставят някои интересни резултати свързани с техния спектрален анализ.

Anubis е блок шифър, който е участник в проекта NESSIE [120]. C-кутията в Anubis е конструирана посредством инволюции. Оказва се, че този вид конструкция лесно може да бъде установена със спектрален анализ тип $\S^{-x,x}$.

CLEFIA е 128 битов блок шифър поддържащ различни видове дължини на ключа от 128, 192 и 256 бита [130]. След спектралния анализ на C-кутиите

установихме редица аномалии. Такива например са вертикалните и хоризонтални червени линии разположени непосредствено до абсцисата и ординатата. Видим е и плътен червен квадрат в най-горния ляв ъгъл на спектралния канал.

Алгоритъмът CMEA е блок шифър, който се е използвал за подсигуряването на мобилните комуникации в САЩ [119]. След анализ на спектралния канал $\S^{0}_{CMEA}$ ясно се очертават (непосредствено до ординатата) хоризонтални червени линии.

Crypton е 128 битов блок шифър участник в AES. C-кутията от неговата първа версия (S_0) [87] е в последствие заменена с 4 C-кутии (S_1, S_2, S_3 и S_4) [88]. Аномалии бяха намерени в S_0 посредством ограничителното множество $\{-8, 8\}$. Във всички ревизирани C-кутии на Crypton също бяха намерени аномалии.

Друг участник в проекта NESSIE е блок шифъра CS-cipher [137]. След спектрален анализ $\S^{0}_{CS}$ се появява интересна графика, с ясно видими модели.

(CSS) [12] се използва за кодирането на DVD дискове. Там също бяха открити аномалии. Допълнително бяха анализирани и публикуваните C-кутии за Enocoro [141], Fantomas [63], FLY [77], Fox [76] и Iceberg [136]. Аномалиите намерени в Enocoro са ясно видими използвайки ограничително множество $E = \{0\}$. Бял правоъгълник покриващ ниските стойности от абсцисата при анализа на Fantomas бяха засечени за спектрален канал $\S^{-4,4}_{Fantomas}$, а по-малки, почти завършени правоъгълници са видими при анализа на FLY за спектрален канал $\S^{-8,8}_{FLY}$. Анализирането на Fox посредством спектрален канал $\S^{-4,4}_{Fox}$ визуализира решетъчноподобна структура. C-кутията в Iceberg е инволюция.

Аномалии бяха открити в Iraqi [143], iScream [64], Khazad [9], Lilliput [3] и Picaro [116]. Спектралният канал $\S^{-1,1}_{Iraqi}$ се различава от случайно генерирани визуализации по лентоподобните структури. По спектралният анализ може да се определи и дали анализираната C-кутия е биективна или не, както е в случая с $\S^{-1,1}_{Iraqi}$. Фракталоподобна структура е видима за спектрален канал $\S^{-4,4}_{iScream}$, докато инволюцията заложена в Khazad лесно се разпознава за $\S^{0}_{Khazad}$. Тетрисоподобна структура се различава за спектрален канал $\S^{-4,4}_{Lilliput}$, докато мрежеста структура е ясно видима за спектрален канал $\S^{-8,8}_{Picaro}$.

Аналогично се очертават и аномалии в Safer [95], Scream [29], SKINNY [5], SNOW 3G [110] и Twofish [127]. В $\S^{-4,4}_{Scream}$ и $\S^{-4,4}_{SKINNY}$ изпъкват любопитни шарки и модели, докато спектрален канал $\S^{-2,2}_{SNOW3G}$ е напълно безцветен. И в двете C-кутии заложени в Twofish, π_0 и π_1, бяха открити аномалии.

Накрая, бяха анализирани C-кутиите заложени в Whirlpool [11], Zorro [53] и ZUC [145]. И в трите бяха открити аномалии в спектрални канали $\S^{0}_{Whirlpool}$, $\S^{-2,2}_{Zorro}$ и $\S^{-8,8}_{ZUCS_0}$.

3.3 Автоматичен спектрален анализ на C-кутии

Процесът по спектрален анализ може да се автоматизира. Нещо повече, както се демонстрира в настоящия раздел, този подход е приложим и за други таблици освен LAT, като DDT, ACT и XOR таблици.

Таблицата S_{LAT} има 2^n колони. Нека бележим с $S^T_{LAT}[i]$ i-та колона на S_{LAT}. Също така ще бележим с $\sigma(S,i,e)$ общия брой срещания на стойностите e и $-e$ в $S^T_{LAT}[i]$, докато с $\sigma_{ind}(S,e)$ ще бележим множеството от индекси на колони от S_{LAT}, такива че:

$$\forall_{i_1 \neq i_2, i_1, i_2 \in \sigma_{ind}(S,e)} : \sigma(S,i_1,e) \equiv \sigma(S,i_2,e).$$

В контекста на случайно генерираните C-кутии, за някоя ограничителна стойност t и две различни стойности e_1 и e_2, $\sigma_{ind}(S,e_1) \equiv \sigma_{ind}(S,e_2)$, където $\sigma(S,i,e_1) \geq t$ и $\sigma(S,i,e_2) \geq t$, е малко вероятно събитие. По време на нашите експерименти, бяха генерирани повече от 10^5 случайни C-кутии. Само в 0.3% от тях бяха открити такива колизии и винаги с дължина 8. Нека бележим една такава подобна колизия с $\Gamma(S,t,e_1,e_2,I)$, където I е множеството от индекси на колони от S_{LAT}.

Такава колизия беше открита в Kuznyechik, която не беше открита като аномалия по време на спектралния анализ. Индексите описващи тази колизия потвърждават наблюденията извършени в [20]. Прилагайки същата стратегия за автоматично търсене по редовете на LAT (вместо по колоните), такива колизии бяха намерени и в други C-кутии, например в тази описана в BelT [124], националния стандарт на Република Беларус. Същият подход е успешно приложим и с DDT таблицата. Така беше открита и колизия в C-кутия π_3 на украинският криптографски алгоритъм Kalyna [109]. Бяха открити множество други колизии при анализа на DDT по редове, както и по ACT (по редове и колони).

Бяха анализирани и спектралните XORT канали на различни популярни C-кутии. Трябва да се отбележи, че визуалната интерпретация на XORT таблицата зависи от начина по който са подредени колоните й - в оригиналната дефиниция лексикографски. Разменяйки обаче тази подредба и сортирайки колоните по поредно участие на координатите във всички възможни линейни комбинации (първо координатите, после двукомпонентните линейни комбинации от координати и т.н; последната колона е линейната комбинация на всички координати), получаваме различна визуална интерпретация. След анализ на C-кутията в BelT посредством оригиналната XORT таблица не бяха открити аномалии. При разместената XORT таблица обаче, бяха открити такива.

Глава 4

Двоични редици и тяхната автокорелация

Редици с ниска автокорелационна функция са важни от практическа гледна точка за редица приложения свързани с обработка на сигналите и информацията. Например, такива редици се използват за достигане на висока резолюция при радари и сонари. Различните изместванията на редици с ниска автокорелация служат за по-добра синхронизация или за идентифициране на потребителите в многопотребителски системи. Поради важността им, тези редици са широко изследвани и в литературата са известни разнообразни методи за построяването им.

Нека $B = (b_0, b_1, \cdots, b_{n-1})$ е двоична редица с дължина $n > 1$, където $b_i \in \{-1, 1\}, 0 \leq i \leq n-1$. Под апериодична автокорелационна функция на B, или AACF, разбираме

$$C_u(B) = \sum_{j=0}^{n-u-1} b_j b_{j+u}, \quad for\ u \in \{0, 1, \cdots, n-1\}.$$

Ще отбележим, че AACF е първоначално дефинирана в интервала

$$\{-n+1, -n+2, \cdots, -2, -1, 0, 1, 2, \cdots, n-1\}.$$

Тъй като AACF е четна функция, $C_u(B) = -C_u(B)$, то можем да я разглеждаме и само в интервала $\{0, 1, \cdots, n-1\}$. $C_0(B)$ се нарича главен страничен лист, а останалите $C_u(B)$ за $u \in \{1, \cdots, n-1\}$ се наричат странични листи. Дефинираме пиковият страничен лист, или PSL, [139] на B така

$$PSL(B) = \max_{0<u<n} |C_u(B)|.$$

Стойността PSL може да бъде представена и в децибели

$$PSL_{db}(B) = 20\log\left(\frac{PSL(B)}{n}\right).$$

Друга важна характеристика на AACF е качествения фактор, или MF, който ни дава съотношението между енергията на главния страничен лист към енергията на страничните листове, т.е.

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

Двоичните редици с ниска автокорелационна стойност могат да бъдат извлечени от различни конструкции като кодовете на Barker [8], m-редици [62], кодове на Gold [61], кодове на Kasami [78], редици на Weil [123], множества на Legendre и други (виж [86][132]).

Редиците на Barker имат най-добри автокорелационни свойства, но най-дългата такава редица е с дължина 13. От друга страна, m-редиците, кодовете на Gold и редиците на Kasami имат идеална периодична автокорелация, но нямат ограничения по страничните листове на тяхната апериодична автокорелация. Както е показано в [104], през годините са предложени различни аналитични конструкции, а чрез компютърно търсене може да се достигне до различни стратегии за генериране на двоични редици със сравнително малки PSL стойности. Чрез пълно претърсване са открити оптималните стойности на PSL за редици с дължина $n \leq 40$[90], $n \leq 48$[7], $n = 64$[34], $n \leq 68$[82], $n \leq 74$[84], $n \leq 80$ [85], $n \leq 82$ [83] и $n \leq 84$ [81]. Почти оптимални стойности за PSL за редици с дължини $85 \leq n \leq 105$ са публикувани в [105], а за $n \geq 106$ в [49].

4.1 Ефикасно генериране на редици с ниска автокорелационна стойност

В този раздел е предложен олекотен и лесен откъм имплементация евристичен алгоритъм за конструиране на двоични редици с рекордни PSL стойности. Такива бяха намерени за голяма част двоични редици с дължини между 106 и 300, с което бяха подобрени голяма част от резултатите публикувани в [49]. Алгоритъмът може да се използва и за конструиране на редици с дължини по-големи от 300.

Тъй като нашата цел е да намалим PSL стойността на дадена редица B, т.е. да минимизираме PSL(B), логично е да се подходи със стратегия за минимизиране на всяка една от стойностите $C_u(B)$, за $u \in \{1, \cdots, n-1\}$. За целта дефинираме следната оценяща функция:

$$F(B) = \sum_{u=1}^{n-1} |C_u(B)|^P = \sum_{u=1}^{n-1} \left(|\sum_{j=0}^{n-u-1} b_j b_{j+u}| \right)^P,$$

където P е магнитуда на оценящата функция, т.е. колкото по-голяма е стойността на P, толкова по-нетолерантна ще е оценящата функция към големи стойности на $C_u(B)$. Извършените експерименти за различни стойности на P в интервала $[3,5]$ ни позволи да изберем най-продуктивните магнитуди. Ниските стойности на P правят оценящата функция твърде толерантна за високите абсолютни стойности на $C_u(B)$, докато високите такива насищат евристичното топологично поле с локални минимуми. Като резултат, за P беше фиксирана стойност 4.

Нека бележим i-та позиция на двоичната редица B с дължина n с b_i. Променяйки i-та позиция на B означава да подменим стойността b_i с $-b_i$. Под околност на B, означена с $N(B)$, ще разбираме множеството от всички двоични редици построени от B на разстояние точно една променена позиция.

Оптимизационният процес взема като вход дължината на проблема n, оценяващата функция F, ограничителна стойност t, две целочислени стойности h_{min} и h_{max}, определящи броя на възможните променени позиции, както и целевата променлива G.

В началото, генерираме случайна двоична редица B с дължина n. След това претърсваме околността на B, търсейки по-добър кандидат, т.е. кандидат с по-малка оценка. Ако някой съсед X (редица от околността) на B има PSL стойност по-малка или равна на G, ние извеждаме като изход X и прекратяваме процеса. Ако обаче няма по-добър съсед в околността на B, то ние сме попаднали в локален минимум B'. За да избегнем безкраен цикъл по претърсване, променяме произволни h на брой елементи от B', където $h \in [h_{min}, h_{max}]$. Такава промяна ще наричаме "трус". Когато t последователни труса не са достатъчни за преодоляване на локалния минимум, ние започваме целия процес отначало. Алгоритъмът приключва, когато двоична редица със зададената PSL стойност или когато заложения лимит от рестартирания е достигнат.

Като краен резултат бяха подобрени PSL стойностите на 95 от общо 195 тествани дължини. В останалите 100 случая бяха достигнати PSL стойности не по-лоши от тези известни в литературата.

Алгоритъмът е подходящ за паралелизиране. Имплементацията му се осъществи посредством езика Python на среден клас компютър с осемядрен процесор. По време на нашите експерименти беше установено, че за достигането на рекордна стойност са нужни от няколко минути до няколко часа.

4.2 Генериране на дълги редици с рекордни PSL стойности

M-редиците, кодовете на Gold, както и редиците на Kasami имат идеални периодични автокорелационни функции, но нямат ограничения свързани с техните странични листи на апериодичните си автокорелационни функции, т.е. техните PSL стойности са непредвидими. Същото се отнася за множествата на Legendre и редиците на Rudin-Shapiro. Дори е трудно да се установи как точно се изменят тези стойности в зависимост от дължината. Съществува предположение, че PSL стойностите на m-редиците нараства с $\mathcal{O}(\sqrt{n})$, което ги прави един от най-предпочитаните методи за директно конструиране на такива редици. Но както е отбелязано в [75]:

> Твърдението, че PSL стойностите на m-редиците нараства с $\mathcal{O}(\sqrt{n})$, което често се среща в радарната литература, е недоказано и все още непотвърдено от данните с които разполагаме. [1]

Както е обобщено в [104], през годините са предложени различни аналитични и компютърни методи за конструирането на двоични редици със малки PSL стойности. Методите CAN [67], ITROX [133], MWISL-Diag, MM-PSL [134] или DPM [80], могат да достигнат до ако не по-добри, то със сигурност не по-лоши стойности от стойностите получени чрез директно прилагане на известните алгебрични конструкции. Трябва да се отбележи обаче, че при по-големи дължини на входната редица, трудността при прилагането на евристичните алгоритми значително нараства. Както е споменато в [102]:

[1] The claim that the PSL of m-sequences grows like $\mathcal{O}(\sqrt{n})$, which appears frequently in the radar literature, is concluded to be unproven and not currently supported by data.

> Като индикация за ефикасността на нашия EA[2], необходимото изчислително време е 58009 секунди или 16.1136 часа за L=1019. За дължини до 4096, необходимото изчислително време нараства квадратично с L. [3]

Главната мотивация на този раздел е създаването на олекотен и ефикасен алгоритъм за евристично генериране на дълги двоични редици с ниска PSL стойност.

Нека означим $C_{n-i-1}(B)$ с $\hat{C}_i(B)$. Тъй като това е просто преподреждане на страничните листи на B, то:

$$B_{PSL} = \max_{0<u<n} |C_u(B)| = \max_{0\leq u<n-1} |\hat{C}_u(B)|.$$

Ще бележим с Ω_Ψ списъка от всички последователни странични листи на Ψ. Изчислението на Ω_Ψ, за някоя случайна редица Ψ, има нелинейна времева сложност. Времевата сложност на тривиалния подход е $\mathcal{O}(n^2)$ (два вложени for цикли). Както е показано в теоремата на Wiener–Khinchin–Einstein [142], тази сложност може да бъде понижена. Въпреки достигнатата времева сложност от $\mathcal{O}(n \log n)$, изискуемата памет драстично нараства.

В този раздел представяме алгоритъм, който изчислява Ω_{Ψ_f} (Ψ_f кореспондира на Ψ с променен бит на позиция f), при съхраняването на Ω_Ψ в паметта, с времева сложност $\mathcal{O}(n)$ и изискуема памет n.

Алгоритъмът е реализиран на езика C на среден клас компютър. По време на нашите експерименти, бяха достигани рекордни PSL стойности за по-малко от секунда. Например, както е показано в [102], необходимото време за достигане на PSL стойност 26, за двоична редица с дължина 1019, е 58009 секунди или 16.1136 часа. За сравнение, предложеният от нас алгоритъм достига тази стойност за по-малко от секунда.

Резултатите от алгоритъма могат да бъдат сравнени с други известни алгоритми, които ще реферираме като колекция A, като CAN [67], ITROX [133], MWISL-Diag, MM-PSL [134], DPM [80], 1bCAN [89]. За целта фиксираме двоични редици с дължини x^2 за $x \in [18,44]$. Разликите между предложения алгоритъм и алгоритмите от колекция A са многобройни. Например, ние не използваме конвертиращи функции, регулярни или квадратични оптимизационни проблеми, както и аритметика с числа с плаваща запетая. Всеки един от алгоритмите беше

[2] EA е съкращение на Evolutionary Algorithm

[3] As an indication of the runtime complexity of our EA[4], the computing time is 58009 s or 16.1136 h for L=1019. For lengths up to 4096, the computing time required empirically shows a seemingly quadratic growth with L.

стартиран 12 пъти, като на предложения от нас алгоритъм беше заложено едноминутно ограничение за работа. Въпреки това, резултатите показаха значително превъзходство на алгоритъма ни над тези от колекция A.

Трябва да се отбележи, че в експериментите проведени в публикациите от колекция A, липсват резултати за двоични редици с дължини по-големи от 2^{12}. Това вероятно се дължи на квадратичното нарастване на сложността им при нарастване на дължината. В предложения от нас алгоритъм обаче, не съществува такъв проблем. За демонстрация сравнихме предложения от нас алгоритъм с резултатите известни след пълното претърсване на m-редиците за някои дължини. И в тези случаи, алгоритъма достигна значително по-добри (рекордни) резултати.

4.3 Хибридни конструкции

M-редицата $M = (x_0, x_1, \cdots, x_{2^m-2})$ с дължина $2^m - 1$ е дефинирана така:

$$x_i = (-1)^{Tr(\beta\alpha^i)}, \text{ for } 0 \le i < 2^m - 1,$$

където α е примитивен елемент от полето $\mathbb{F}_{2^m}$, $\beta \in \mathbb{F}_{2^m}$, и Tr е трейс функцията от $\mathbb{F}_{2^m}$ в $\mathbb{F}_2$.

При p просто число, редицата на Legendre L с дължина p е дефинирана така:

$$L_i = \begin{cases} 1, \text{ ако } i \text{ е квадратичен остатък по mod } p \\ -1, \text{ иначе.} \end{cases}$$

Бележим с $B \leftarrow \rho$ двоичната редица получена от B, прилагайки ρ последователни измествания в ляво. По дефиниция, $B \leftarrow |B| \equiv B$. Ако b_i е елемент от B на позиция i, ще бележим с $b_i^{\leftarrow\rho}$ елемента от $B \leftarrow \rho$ на позиция i.

Както показахме в Раздел 4.2 или в [44], въпреките линейните сложности заложени в алгоритъма, процесът по PSL-оптимизация си остава трудна задача за по-големи дължини. След извършен детайлен анализ, предложените от нас алгоритми бяха ревизирани за ефикасно генериране на двоични редици с ниски PSL стойности при значителни по-големи дължини.

По време на експериментите беше използван среден клас шестядрен компютър. След направените корекции за малки дължини, предложения от нас алгоритъм достигна до всички известни оптимални PSL стойности, т.е. дължини в интерва-

ла [1,82]. В повечето случаи, за достигането на оптимална стойност за дадена дължина беше нужно по-малко от минута.

Бяха проведени и експерименти за дължини до 300. Почти всички известни в литературата резултати бяха подобрени. По-точно, бяха подобрени 179 от 195 случая.

В [35], са публикувани най-добрите резултати постигнати посредством квантовия компютър D-Wave 2, като за двоична редица с дължина 128 е достигната PSL стойност 8. За сравнение, предложения от нас алгоритъм достигна PSL стойност 6. За по-големи дължини, например 256, най-добрата стойност достигната от D-Wave 2 е 12, докато ние достигнахме до 10. Като последен пример, стартирахме предложения от нас алгоритъм за дължина 426. Той достигна PSL стойност 17, най-добрата стойност достигната от квантовия компютър, за по-малко от секунда. Нещо повече, предложения от нас алгоритъм достигна до стойности 16 и 15, отново в рамките на секунда. За достигането на нова рекордна PSL стойност от 14 обаче бяха нужни 199 секунди.

Наскоро, в [36] е предложен многонишков еволюционен алгоритъм. След проведените експерименти, успяхме да подобрим почти всички резултати от гореспоменатата статия - обикновено за по-малко от секунда. Например, най-добрата PSL стойност за редица с дължина 3000 публикувана в [36] е 51. Алгоритъмът предложен в този раздел достига стойности от 44 и 43 съответно за 111 и 371 секунди.

Кога една редица се определя като къса или дълга е неизяснен въпрос. Тази характеристика е свързана по-скоро с възможностите на дадения алгоритъм, отколкото с дължината сама по себе си. На практика, някои алгоритми дори няма да могат да бъдат стартирани при големи стойности. Например, както е споменато в [35], квантов компютър D-Wave 2 с 512 кюбита е ограничен откъм обработка на редици с дължини до 426. Един 2048 кюбитов квантов компютър D-Wave 2 ще бъде ограничен до двоични редици с приблизителна дължина 2000.

За удобство, алгоритъмът предложен в този раздел ще бележим с $\mathscr{A}$, като сме фиксирали магнитуда на оценящата фукция на 4.

4.3.1 $\mathscr{A}$ като разширение на m-редиците

Предложена е следната конструкция:

- Избираме примитивен полином f от F_{2^m}.
- Фиксираме елемент a над F_{2^m}.

- Представяме f чрез шифт-регистър с линейна обратна връзка $\mathscr{L}$.
- Разширяваме $\mathscr{L}$ до двоична редица L, $|L| = 2^m - 1$.
- Стартираме $\mathscr{A}$ с вход L.

Примитивните полиноми над F_{2^m} могат да бъдат изчислени предварително. PSL стойността на L може да бъде избрана така, че да има възможно най-малка стойност.

Използвайки този подход, може да достигнем до рекордно ниски PSL стойности. Например за дължина 131071 достигнахме стойност 359 за по-малко от 2 минути. След около 46 минути алгоритъмът достигна стойност 356. За съжаление, тази процедура е приложима единствено за редици с дължини $2^n - 1$. В следващия раздел разширяваме този подход за дължини p, където p е просто число.

4.3.2 $\mathscr{A}$ като разширение на редиците на Legendre

Предложена е следната конструкция:

- Избираме просто число p.
- Построяваме редицата $L = [t_1, t_2, \cdots, t_p]$.
- За i, такова че $i \in N$, $1 \leq i \leq p$, и в случай че i е квадратичен остатък след деление с p, заменяме t_i с 1. Иначе, заменяме t_i с -1.
- Стартираме $\mathscr{A}$ с вход L.

Както показва наличната информация публикувана в [42], малко вероятно е редица на Legendre с дължина p, където $p > 235723$, или която и да е нейна ротация, да достигне до PSL стойност по-малка от $\sqrt{p}$. Имайки това предвид, бяха стартирани експерименти с алгоритъм $\mathscr{A}$ (магнитуд 8), с вход ротирана редица на Legendre с дължина 235747 (следващото просто число след 235723). Използвайки ротация 60547, беше открита двоична редица с PSL стойност 508. След това $\mathscr{A}$ успя значително да подобри този резултат. За по-малко от 25 минути, използвайки само 1 нишка от процесора модел Xeon-2640 с базова чистота 2.50 GHz, беше открита двоична редица с PSL стойност 408.

Тъй като $\sqrt{235747} \approx 485.54$, се вижда че 408 е значително по-малка от очакваната стойност 485.54. Оставайки $\mathscr{A}$ за около 2 часа, беше достигната и PSL стойност 400 (виж [43]).

4.3.3 Апериодична автокорелация на ротирани двоични редици

Както е показано в [47], детайлното проучване на m-редиците може да ни даде важни закономерности и взаимовръзки свързани с определянето на апериодичната автокорелационна характеристика. Намирането на PSL оптимални m-редици обаче е сложна и времеемка процедура - на всяка итерация, освен PSL стойността на дадена двоична редица B, ние трябва да изчислим и PSL стойността на всички възможни нейни ротации. В [75], след пълно обхождане, са публикувани всички PSL оптимални редици с дължини до $2^{15}-1$. По-късно, в [47], са публикувани и такива за дължини $2^{16}-1$ и $2^{17}-1$.

Подобно на изчислителния проблем свързан с m-редиците, намирането на оптимални редици на Legendre, с или без ротации, също представлява голямо предизвикателство. В [128], Фиг.4, е публикуван списък на оптималните PSL редици на Legendre само до дължини 3500.

Процедурата по намиране на минималната PSL стойност измежду всички ротации на дадена двоична редица е най-времеемката част от тези изчисления. В този раздел е описано как тази процедура може да се имплементира като перфектно балансиран паралелен алгоритъм, особено подходящ за процесори с видео обработка. С такъв алгоритъм бяха конструирани всички оптимални m-редици, с или без ротация, за дължини $2^{18}-1$, $2^{19}-1$ и $2^{20}-1$, както и всички оптимални редици на Legendre, с или без ротация, за дължини до 432100.

Теорема 4.3.1. Нека имаме двоична редица $B = b_0 b_1 \cdots b_{n-1}$ с дължина n. Тогава е изпълнено следното свойство:

$$\hat{C}_i(B \leftarrow 1) - \hat{C}_i(B) = b_0 \left(b_{i+1} - b_{n-i-1}\right).$$

Теорема 4.3.2. Нека имаме двоична редица $B = b_0 b_1 \cdots b_{n-1}$ с дължина n. Тогава, разликата $\hat{C}_i(B \leftarrow \rho) - \hat{C}_i(B \leftarrow (\rho-1))$ е равна на $b_{(\rho-1) \bmod n}\left(b_{(i+\rho) \bmod n} - b_{(n-i+\rho-2) \bmod n}\right)$.

Нека с Ω_B означим списъка от всички странични листи на дадена двоична редица B с дължина n, или: $\Omega_B = \left[\hat{C}_0(B), \hat{C}_1(B), \cdots, \hat{C}_{n-2}(B)\right]$. След използването на Теорема 4.3.2 и връзката между $\Omega_{B\leftarrow\rho}$ и $\Omega_{B\leftarrow(\rho-1)}$, можем да изчислим $\Omega_{B\leftarrow\rho}$, при дадено $\Omega_{B\leftarrow(\rho-1)}$, използвайки $n-1$ различни паралелни нишки. Две важни свойства:

- Нишките са независими една от друга.

- Нишките са перфектно балансирани от гледна точка на синхронизация, т.е. за две различни нишки t_i и t_j, броят на необходимите аритметични операции за завършване на изчислителния процес на t_i и t_j е еднакъв.

Този сценарий е напълно съвместим с модела SIMD [51]. Можем да делегираме изчислението на $\hat{C}_i(B \leftarrow \rho)$ единствено на нишка t_i, тъй като то е независимо от изчислението на другите ротации. Нещо повече, за допълнителна оптимизация на изчислителния процес, можем последователно да презаписваме стойностите на $\hat{C}_i(B)$, т.е. Ω_B с последващите ротации $\Omega_{B\leftarrow\rho}$, за $\rho \in [1, n-1]$.

Реализацията на алгоритъма беше осъществена посредством няколко програмни езика [5]. За целта бяха използвани и видео ускорители.

За сравнение, ако използваме предложения алгоритъм във видео ускорител с приблизително 1200 ядра, заедно със среден клас компютър с 6 ядра, необходимото време за намиране на PSL-оптималната редица измежду дадена редица B с дължина $2^{20}-1$ и всички възможни ротации на B ще е приблизително 191 секунди. За извършването на същото изчисление посредством популярната библиотека NumPy [108] ще са нужни 36 години. Ускорението е приблизително $2^{22.5}$.

С предложения алгоритъм бяха намерени всички PSL-оптимални m-редици с дължини $2^{18}-1$, $2^{19}-1$ и $2^{20}-1$, както и всички PSL-оптимални редици на Legendre до дължини 432100. Данните от проучването водят до предположението, че не съществува редица на Legendre L с дължина n по-голяма от 235723, такава че L или която и да е ротация на L имат PSL стойност по-малка или равна на $\sqrt{n}$.

[5] C, Python, SageMath, CUDA

Глава 5

Двоични редици и техният качествен фактор

Проблемът с качествения фактор, или MF, на двоичните редици е важен за голям брой дисциплини, като цифровите комуникации, радари, системи за модулации, машинно тестване, теория на информацията, физика и химия. MF проблемът е често реферирaн като един от най-трудните оптимизационни проблеми и дори съществува предположение, че стохастични конструкции няма никога да достигнат MF стойности над 5 за дълги двоични редици (редици с дължини над 200). В настоящия раздел се разглеждат няколко интересни математически свойства на изкривено-симетричните двоични редици при промяната на точно два елемента в тях. Използвайки тези свойства, изискуемата памет на съвременните алгоритми за търсене на двоични редици с ниски MF стойности може да бъде редуцирана от n^2 до n. Например, предложения в този раздел алгоритъм може да оптимизира случайно генерирани изкривено-симетрични двоични редици с големи дължини (до $10^5 + 1$) до изкривено-симетрични редици с MF стойност над 5.

5.1 Качественият фактор на изкривено-симетричните двоични редици

Ако с F_n бележим оптималната (най-голямата) стойност на качествения фактор измежду всички двоични редици с дължина n, то MF-проблема може да бъде представен като намиране на стойността $\limsup_{n\to\infty} F_n$. Съществуват няколко предположения за $\limsup_{n\to\infty} F_n$. Първото предположение, публикувано в [69], прогнозира $\limsup_{n\to\infty} F_n = 6$. По-крайното предположение $\limsup_{n\to\infty} F_n = \infty$

е дадено от Littlewood [91]. В [27], $\limsup_{n\to\infty} F_n = 5$. Golay [58] предполага, че $\limsup_{n\to\infty} F_n$ е много близко до 12.32. Въпреки това, в [59] той добавя че "...няма да се намери систематичен анализ, който да позволи построяването на редици с качествен фактор над 6 ... [1]". В [21] се предполага, че $\limsup_{n\to\infty} F_n > 6.34$.

През последните 50 години, освен теоретичните атаки, са събрани и множество доказателства и данни за този проблем, посредством два основни метода - пълно обхождане и евристични алгоритми.

Оптималните MF стойности на редици с дължини $n \leq 60$ са публикувани в [98]. Двадесет години по-късно, списъкът е разширен до $n \leq 66$ [111]. Най-големите известни стойности за F_n са 14.1 и 12.1 за n съответно 13 и 11. Ще отбележим, че и двете споменати двоични редици са съставени от редиците на Barker [8]. Всъщност, в [74] авторът публикува няколко предизвикателства свързани с MF проблема. Първото предложено предизвикателство е да се намери двоична редица X с дължина $n > 13$, за която $F(X) \geq 10$.

Един разумен подход за намиране на двоични редици с почти оптимална MF стойност е посредством рестрикция върху структурата на множеството от редици. Една такава рестрикция, предложена от Golay [55], е множеството на изкривено-симетричните редици. За редицата $(b_0, b_1, \cdots, b_{2l})$ с нечетна дължина $n = 2l + 1$, рестрикцията се дефинира така: $b_{l+i} = (-1)^i b_{l-i}$ за $i = 1, 2, \cdots, l$.

Golay установява, че нечетните редици на Barker са изкривено-симетрични [57]. Както е показано и в [55], всички странични листи на изкривено-симетричните редици с четни индекси са равни на 0.

Оптималните MF стойности на всички изкривено-симетрични редици с нечетни дължини $n \leq 59$ са открити от самия Golay [57]. По-късно, и за дължините $n \leq 69$ и $n \leq 71$ съответно в [60] и [39], а за дължините $n \leq 89$ и $n \leq 119$ съответно в [118] и [111].

Проблемът по минимизирането на F_n е известен и като LABS проблем. Той е добре известен в теоретичната физика и химията. MF проблемът е атакуван от различни алгоритми, като този предложен в [111], стохастичния алгоритъм предложен в [66], хибридния алгоритъм в [52], и като тези предложени в [39][99]. Тъй като големината на пространството на търсене нараства подобно на 2^n, трудността по намирането на дълги редици с почти оптимална F_n стойност осезаемо се увеличава. В [39] проблемът е описан като "един от най-трудните оптимизационни проблеми[2]".

[1] ...no systematic synthesis will ever be found which will yield higher merit factors [than 6]...

[2] ... amongst the most difficult optimization problems...

Основният подход заложен в евристичните алгоритми е търсене в пространството от решения с малки промени на текущия кандидат. При наличие на изкривено-симетрична рестрикция, Golay препоръчва [56] това търсене да се извършва посредством промяната на един или два елемента. Ако новият кандидат притежава по-добра MF стойност, оптимизационният процес го приема като текущ кандидат, продължавайки процеса от него. Съществуват различни стратегии за случаите, когато в околността на текущия кандидат няма по-добри такива.

Най-добрите резултати свързани с изкривено-симетричните редици са публикувани в [52], [23], [25] и [26]. В [52], авторите предлагат алгоритъм с ефикасен метод за изчисление на характеристиките на дадена двоична редица L', такава че L' е на точно една промяна от L, и предполагайки, че няколко умножения на елементи от L са предварително записани в паметта. По-точно, квадратна $(n-1, n-1)$ матрица $\tau(S)$, такава че $\tau(S)_{ij} = s_j s_{i+j}$ за $j \leq n-i$. По-късно, в [23] е предложен принципът на непресичащото се търсене [93]. Използвайки графи на Hasse авторите демонстрират, че често съвременните алгоритми атакуващи MF проблема попадат в цикли. За да избегнат този негативен ефект, те предлагат използването на хеш таблици. В [25] е предложен алгоритъма xLastovka. Появява се и концепцията на приоритизиращата опашка. Рекордни стойности на изкривено-симетрични двоични редици с дължини между 301 и 401 са публикувани в [26].

Гореспоменатите алгоритми разчитат на $\tau(S)$ матрицата. Тя значително ускорява скоростта на оценяването на съседи на кандидата, достигайки времева сложност $O(n)$. Нейния недостатък обаче е, че изискуемата памет нараства квадратично. Поради тази причина, съвременните алгоритми не могат да бъдат стартирани (поради хардуерни ограничения) за двоични редици с големи дължини.

В този раздел се предлага алтернатива на $\tau(S)$ матрицата, която свежда квадратичната изискуема памет до линейна такава. Предложената оптимизация е лесна за реализиране. Например, в [22] е публикувана колекция от рекордните MF стойности на изкривено-симетрични двоични редици с дължини от 5 до 449. За 449 рекордната стойност е 6.5218. Използвайки нисък клас процесор Xeon-2640 с честота от 2.50 GHz, и чрез имплементиране на предложената от нас оптимизация в lssOrel алгоритъма [22], успяхме да достигнем до изкривено-симетрична редица за същата дължина с рекордна MF стойност 6.5319. Необходимото време беше около 24 часа. За сравнение, предишният рекорд е постигнат посредством SLING инфраструктурата (100 процесора) за 4 дни.

В този раздел се предлага и алгоритъм, който има минимални изисквания за памет. Това може да бъде полезно при имплементацията му в многонишкови

системи или видео ускорители. По време на проведените експерименти, с предложения алгоритъм бяха достигнати MF стойности строго по-големи от 5, за двоични редици с различни дължини до $10^5 + 1$. От това следва, че предположението на Bernasconi е доста песимистично.

Нека вземем изкривено-симетрична редица описана като списък $L = [b_0, b_1, \cdots, b_{n-1}]$ с нечетна дължина $n = 2l + 1$. Ако списъкът от странични листи на L обозначим с W, тогава:

$$W = [C_{n-1}(L), C_{n-2}(L), \cdots, C_1(L), C_0(L)],$$

където

$$C_u(L) = \sum_{j=0}^{n-u-1} b_j b_{j+u}, \quad for\ u \in \{0, 1, \cdots, n-1\}.$$

За удобство, нека обозначим обърнатия на обратно W списък с S, т.е.:

$$S = [\hat{C}_0(L), \hat{C}_1(L), \cdots, \hat{C}_{n-2}(L), \hat{C}_{n-1}(L)],$$

където $\hat{C}_{n-i-1}(L) = C_i(L)$, за $i \in \{0, 1, \cdots, n-1\}$. Тогава,

$$\hat{C}_i(L) = C_{n-i-1}(L) = \sum_{j=0}^{n-(n-i-1)-1} b_j b_{j+(n-i-1)}.$$

Следователно,

$$\hat{C}_i(L) = \sum_{j=0}^{i} b_j b_{j+n-i-1}, \quad i \in \{0, 1, \cdots, n-1\}.$$

Ще бележим i-тия елемент на даден списък A с $A[i]$. За удобство, първият индекс е 0, не 1. Например,

$$W[n-1] = S[0] = \hat{C}_0[L] = C_{n-1}(L).$$

Тъй като L е изкривено-симетрична двоична редица, тя има следните свойства:

- $S[i] = 0$, за нечетни стойности на i.
- $L[l-i] = (-1)^i L[l+i]$.

Следователно, списъкът от странични листи S може да бъде представен така:

$$S = [\hat{C}_0(L), 0, \hat{C}_2(L), 0, \cdots, 0, \hat{C}_{n-3}(L), 0, \hat{C}_{n-1}(L)].$$

За удобство, ще означаваме с S_i $(i-1)$-тия елемент от S, т.е. $S_i = S[i-1]$.
Тогава, за всяка нечетна стойност r, имаме:

$$S_r = \hat{C}_{r-1}(L) = \sum_{j=0}^{r-1} b_j b_{j+n-r+1-1} = \sum_{j=0}^{r-1} b_j b_{j+n-r} = \sum_{j=1}^{r} b_{j-1} b_{j-1+n-r}.$$

Предното наблюдение може да бъде записано и така:

$$S_r = \sum_{j=1}^{r} b_{j-1} b_{j-1+n-r} = \sum_{i=1}^{r} L[i-1]L[n+i-r-1].$$

От една изкривено-симетрична редица L с дължина $n = 2l+1$, при промяна на елементите на позиции q и $n-q-1$, за някое фиксирано $q \in \{0,1,\cdots,l\}$, се получава редица L^q, която също е изкривено-симетрична. Нека означим списъка от странични листи на L^q с S^q:

$$S^q = \left[\hat{C}_0(L^q), 0, \hat{C}_2(L^q), 0, \cdots, 0, \hat{C}_{n-3}(L^q), 0, \hat{C}_{n-1}(L^q)\right].$$

Теорема 5.1.1. При дадени две изкривено-симетрични редици L и L^q с дължини $n = 2l+1$, и със списъци на странични листи съответно S и S^q, където $q < l$, са изпълнени следните свойства:

I За $\forall e$, такова че e е четно, $S^q_e - S_e = 0$.

II Ако r е нечетно и $r \le q$, то $S^q_r - S_r = 0$.

III Ако r е нечетно и $r > q$, $r < n-q$, и $q \neq r-q-1$, то:

$$S^q_r - S_r = -2\left(L[q]L[n+q-r] + L[r-q-1]L[n-q-1]\right).$$

IV Ако r е нечетно и $r > q$, $r < n-q$, и $q = r-q-1$, то $S^q_r - S_r = 0$.

V Ако r е нечетно и $r \ge n-q$, и $q \neq r-q-1$, то:

$$S^q_r - S_r = -2L[n-q-1]L[2n-q-r-1] - 2L[q+r-n]L[q] -$$

$$-2L[q]L[n+q-r] - 2L[r-q-1]L[n-q-1].$$

VI Ако r е нечетно и $r \ge n-q$, и $q = r-q-1$, то:

$$S^q_r - S_r = -2L[n-q-1]L[2n-q-r-1] - 2L[q+r-n]L[q].$$

Ще отбележим, че Теорема 5.1.1 покрива всички възможни позиции за промяна на бит.

Теорема 5.1.2. При дадени две изкривено-симетрични редици L и L^q с дължини $n = 2l + 1$, където L^q се получава от L с q-ти и $n-q-1$-ви променен бит, за някое фиксирано $q < l$, и със списъци на странични листи съответно S и S^q, е изпълнено следното свойство:

$$\mathbb{E}(L^q) = \mathbb{E}(L) + \sum_{r=q+1, r\neq 2q+1}^{n-q-1} (16 + \sigma\kappa\varepsilon_1) + \sum_{r=n-q, r\neq 2q+1}^{n-1} (\kappa(\varepsilon_2 + \sigma\varepsilon_1) + 32 + 32\sigma\varepsilon_1\varepsilon_2) + \\ + \sum_{r\geq n-q, r\leq n-1, r=2q+1} (16 + \kappa\varepsilon_2), \tag{5.1}$$

където $\sigma = (-1)^{l-q}$, $\kappa = -8S_rL[q]$, $\varepsilon_1(r) = L[r-q-1]$, $\varepsilon_2(r) = L[q+r-n]$.

Това свойство ни позволява да редуцираме изискуемата от съвременните алгоритми памет за атакуване на MF проблема от n^2 до n. Например, при използването на само една нишка от процесора, построяването на помощната $\tau(S)$ матрица за редица с дължина 5000 изисква около 95.37 MB, докато посредством предложения алгоритъм, изискуемата памет е 19.53 KB. Нещо повече, от практическа гледна точка, подмяната на $\tau(S)$ матрицата с предложената от нас структура не представлява особена трудност, като запазва времевата сложност на дадения алгоритъм.

5.1.1 За предположението на Bernasconi

Както споменахме, в [13] Bernasconi предполага, че никоя стохастична процедура няма да достигне MF стойност по-висока от 5 за дълги редици (редици с дължина по-голяма от 200). Това предположение е направено през далечната 1987 година. От тогава е изминало доста време и са събрани известен брой доказателства, че все пак такива стохастични алгоритми могат да бъдат създадени. Например, евристични алгоритми, с които могат да се конструират двоични редици с нечетна дължина до около 500 и MF стойност 5. Но предположението на Bernasconi изглежда вярно, ако неговата граница от 200 се осъвремени, предвид наличния прогрес на съвременната изчислителна мощ. Разбира се, ако се открие стохастичен алгоритъм, който може да достига много по-дълги редици с MF стойност по-голяма от 5, то предположението на Bernasconi може да се окаже доста песимистично.

По време на проведените експерименти, използвайки алгоритъм с подобренията описани в предните раздели, успяхме да конструираме изкривено-симетрични редици с дължини до 100001 и MF стойности винаги строго по-големи от 5.

5.1.2 Нови класове двоични редици с високи MF стойности

Въпреки множеството от резултати свързани с изкривено-симетричните двоични редици, двоичните редици с четни дължини са значително по-слабо проучени. Това не е изненадващо, тъй като рестрикцията предложена от Golay е приложима единствено за редици с нечетна дължина.

В този раздел, мотивирани от липсата на рестрикция приложима и за редици с четна дължина, се предлагат нови класове двоични редици.

Дефиниция 5.1.1 (Псевдо-изкривено-симетрични двоични редици). Наричаме дадена редица $P = a||X = Y||b$ псевдо-изкривено-симетрична, ако X или Y са изкривено-симетрични, за някое $a \in \{-1, 1\}$ или $b \in \{-1, 1\}$.

Твърдение 5.1.1. Списъкът от странични листи на псевдо-изкривено-симетричните двоични редици се състои от алтернириращи единици по абсолютна стойност.

Твърдение 5.1.2. При дадена изкривено-симетрична двоична редица $B = (b_0, b_1, \cdots, b_{n-1})$ със списък на страничните листи

$$S_B = \left[\hat{C}_0(B), \hat{C}_1(B), \cdots, \hat{C}_{n-2}(B), \hat{C}_{n-1}(B)\right],$$

е изпълнено следното свойство:

$$\mathbb{E}(P) = \mathbb{E}(B) + n + 2b_n\delta,$$

където P е псевдо-изкривено-симетрична редица $B||b_n$ и $\delta = \sum_{u=0,u_{even}}^{n-2} \hat{C}_u(B)b_{u+1}$.

Това свойство е полезно при преобразуването на даден алгоритъм търсещ изкривено-симетрични редици, обозначен с $\mathscr{A}$, към алгоритъм търсещ псевдо-изкривено-симетрични такива. Без значение от сложността на $\mathscr{A}$, той може да се представи в следния вид $\cdots||\mathbb{L}_1||\cdots||\mathbb{L}_2||\cdots||\mathbb{L}_n||\cdots$, като $\mathbb{L}_i$ обозначават тези етапи на $\mathscr{A}$, в които се обявяват намерените по-добри кандидати, т.е. локалните оптимуми. Можем лесно да заменим $\mathbb{L}_i$ с $\mathbb{L}_i||\mathbb{T}_i$, където $\mathbb{T}_i$ е тривиална процедура с линейна сложност, която изчислява MF стойностите на псевдо-изкривените-симетрични редици $L_i||1$ и $L_i||-1$. Трябва да се отбележи, че новия вид на

алгоритъма, т.е. $\mathscr{B} = \cdots||\mathbb{L}_1||\mathbb{T}_1||\cdots||\mathbb{L}_2||\mathbb{T}_2||\cdots||\mathbb{L}_n||\mathbb{T}_n||\cdots$ не променя по никакъв начин основната функция на $\mathscr{A}$. Нещо повече, тъй като тези проверки се извършват единствено при достигане на локален оптимум, добавената функционалност забавя $\mathscr{A}$ с пренебрежимо малко време.

Можем да разширим функционалността на съществуващите алгоритми и със следното наблюдение.

Твърдение 5.1.3. При дадена изкривено-симетрична двоична редица $B = b_0||B'||b_{n-1}$, двоичните редици $b_0||B'$ и $B'||b_{n-1}$ са псевдо-изкривено-симетрични.

Твърдение 5.1.4. При дадена изкривено-симетрична двоична редица $B = (b_0, b_1, \cdots, b_{n-1}) = b_0||B'||b_{n-1}$ със списък на страничните листи

$$S_B = \left[\hat{C}_0(B), \hat{C}_1(B), \cdots, \hat{C}_{n-2}(B), \hat{C}_{n-1}(B)\right],$$

е изпълнено следното свойство:

$$\mathbb{E}(P) = \mathbb{E}(B) + n - 3 + 2b_{n-1}\delta,$$

където P е псевдо-изкривено-симетричната редица $b_0||B'$ и $\delta = \sum_{u=1,u_{even}}^{n-2} -\hat{C}_u(B)b_u$.

Последното свойство допълнително разширява възможностите на съществуващите алгоритми, т.е. даден алгоритъм $\mathscr{A}$, за търсене на изкривено-симетрични редици с висока MF стойност и с нечетна дължина n, може да бъде лесно трансформиран до алгоритъм $\mathscr{B}$, търсещ едновременно изкривено-симетрични редици с висока MF стойност и с нечетна дължина n, както и псевдо-изкривено-симетрични редици с висока MF стойност и с четни дължини $n-1$ и $n+1$.

Дефиниция 5.1.2 (Рестрикционен клас на двоична редица). Ще наричаме класа съставен от двоични редици с дължина n, с фиксирани първи k елемента, рестрикционен клас от степен k на двоичните редици с дължина n. Ще го бележим с R_n^k. Ако двоичната редица е изкривено-симетрична ще я бележим с $\mathscr{R}_n^k$.

Проблемът с разбиването на числа е добре известен комбинаторен проблем, т.е. по-колко различни начина дадено естествено число n може да се представи като сума от положителни числа, известно още като функцията $p(n)$. Въпреки, че представянето на $p(n)$ в общия случай е отворен въпрос, част от стойностите му могат да бъдат намерени в онлайн енциклопедията на целите числа (OEIS) под номер A000041 [1].

От друга страна, търсенето на изкривено-симетрични двоични редици с дължина n може да бъде паралелизирано на $|\mathscr{R}_n^k|$ части. За да минимизираме броя на паралелните части, трябва да бъдат разгледани следните операции над изкривено-симетричната двоична редица $B = (b_0, b_1, \cdots, b_{n-1})$:

- Обръщане на B дефинирано като оператор δ_1: $\delta_1(B) = (b_{n-1}, \cdots, b_1, b_0)$.
- Допълнение на B дефинирано като оператор δ_2: $\delta_2(B) = (\overline{b_0}, \overline{b_1}, \cdots, \overline{b_{n-1}})$, където $\overline{b_i} = -b_i$.
- Алтерниращо допълнение на B дефинирано като оператор δ_3 : $\delta_3(B) = (\cdots, \overline{b_{i-2}}, b_{i-1}, \overline{b_i}, b_{i+1}, \overline{b_{i+2}}, \cdots)$.

И трите оператора оставят енергията на B непроменена. Ако допълнително включим и оператора идентитет δ_0, то ще получим група G, такава че $ord(G) = 8$. От [111], можем да изведем точна формула за симетричните класове в G с дължина k: $2^{k-3} + 2^{\lfloor \frac{k}{2} \rfloor - 2 + (k \bmod 2)}$. Същата формула се получава по сбора на редовете на триъгълника на Losanitsch (OEIS, sequence A005418 [2]), свързана с неговата работа по парафините [92]. Последните няколко наблюдения ни позволяват да разделим пространството на търсене от $p(k)$ до $2^{k-3} + 2^{\lfloor \frac{k}{2} \rfloor - 2 + (k \bmod 2)}$ на непресичащи се подмножества.

Дефиниция 5.1.3 (Потенциал на рестрикционен подклас). За изкривено-симетрична двоична редица $B = (b_0, b_1, \cdots, b_{n-1})$, фиксираме разбиване k: $t_0, t_1, \cdots, t_g$, такова че $\sum_{i=0}^{g} t_i = k$. Разбиването може да бъде проектирано върху изкривено-симетрична двоична редица със следната процедура:

$$R = \underbrace{a\cdots a}_{t_0}\underbrace{\overline{a}\cdots\overline{a}}_{t_1}\underbrace{a\cdots a}_{t_2}\underbrace{\overline{a}\cdots\overline{a}}_{t_3}\cdots\underbrace{(-1)^g a\cdots(-1)^g a}_{t_g}\underbrace{u_1u_2u_3\cdots u_{n-2k-2}u_{n-2k-1}u_{n-2k}}_{\text{нефиксирани елементи}}\underbrace{f_1f_2f_3\cdots f_{k-2}f_{k-1}f_k}_{\text{фиксирани елементи}}$$

Последните k елемента f_i са също фиксирани поради първите k елемента от редицата и нейното основно свойство на изкривена-симетричност. Също така $a, \overline{a}, (-1)^g a, u_i, f_i \in \{-1, 1\}$. Под потенциал на редицата R ще разбираме енергията на троичната редица R^z, където:

$$R^z = \underbrace{a\cdots a}_{t_0}\underbrace{\overline{a}\cdots\overline{a}}_{t_1}\underbrace{a\cdots a}_{t_2}\underbrace{\overline{a}\cdots\overline{a}}_{t_3}\cdots\underbrace{(-1)^g a\cdots(-1)^g a}_{t_g}\underbrace{000\cdots000}_{n-2k\text{ нули}}\underbrace{f_1f_2f_3\cdots f_{k-2}f_{k-1}f_k}_{\text{фиксирани}}$$

5.1.3 Алгоритъм за намиране на двоични редици с произволна дължина и висока MF стойност

Предложеният алгоритъм е реализиран на C++. По време на експериментите беше използван среден клас компютър с 8 ядра (16 нишки). Бяха подобрени всички известни MF рекорди за дължините в интервала 225-451, изчислени от суперкомпютърна мрежа. Също така, бяха открити псевдо-изкривено-симетрични редици, т.е. редици с четни дължини, за дължини в интервала 225-512 с MF стойности по-големи от 7. За пълнота е публикуван пълен списък с рекорди, за четни и нечетни дължини до 2^8 с MF стойности строго по-големи от 8, както и пълен списък с рекорди, за четни и нечетни дължини до 2^9 с MF стойности строго по-големи от 7.

Като допълнителна демонстрация за ефикасността на предложения алгоритъм, той беше стартиран за оптимизация на двоични редици с дължини 573 и 1009. За тези дължини е установено, че за един от водещите алгоритми са нужни съответно 32 и 46774481153 години за достигането на MF стойност по-голяма от 6.34. Алгоритъмът описан в този раздел достига MF стойности по-големи от 6.34 за тези дължини в рамките на няколко часа.

5.2 Спектрален анализ на C-кутии посредством апериодични автокорелационни функции

В този раздел е предложена нова стратегия за спектрален анализ на C-кутии, третирайки всички $\binom{n}{2}$ колони съставени от двукомпонентни линейни комбинации от координати на дадена C-кутия $S(n,n)$ като двоични редици. Такава стратегия за прилагане на спектрален анализ на получените странични листи има смисъл, тъй като това ще разкрие евентуални корелации измежду компонентите на кутията. Като пример, посредством този подход, бяха открити аномалии в C-кутиите на BelT, CSS, Safer и SKINNY.

Библиография

[1] Oeis a000041. https://oeis.org/A000041. Accessed: 2022-05-30.

[2] Oeis a005418. https://oeis.org/A005418. Accessed: 2022-05-30.

[3] Adomnicai, A., Berger, T. P., Clavier, C., Francq, J., Huynh, P., Lallemand, V., Le Gouguec, K., Minier, M., Reynaud, L., and Thomas, G. (2019). Lilliput-ae: a new lightweight tweakable block cipher for authenticated encryption with associated data. Submitted to NIST Lightweight Project.

[4] Ahmad, M., Bhatia, D., and Hassan, Y. (2015). A novel ant colony optimization based scheme for substitution box design. Procedia Computer Science, 57:572–580.

[5] Andreeva, E., Lallemand, V., Purnal, A., Reyhanitabar, R., Roy, A., and Vizár, D. (2019). Forkae v. Submission to NIST lightweight cryptography project.

[6] Aoki, K., Ichikawa, T., Kanda, M., Matsui, M., Moriai, S., Nakajima, J., and Tokita, T. (2000). Camellia: A 128-bit block cipher suitable for multiple platforms—design andanalysis. In International Workshop on Selected Areas in Cryptography, pages 39–56. Springer.

[7] Baden, J. and Cohen, M. (1990). Optimal peak sidelobe filters for biphase pulse compression. In IEEE International Conference on Radar, pages 249–252. IEEE.

[8] Barker, R. H. and Jackson, W. (1953). Group synchronization of binary digital systems in Communication Theory. Academic Press, New York, pages 273–287.

[9] Barreto, P. and Rijmen, V. (2000). The khazad legacy-level block cipher. Primitive submitted to NESSIE, 97:106.

[10] Barreto, P., Rijmen, V., et al. (2000a). The Whirlpool hashing function. In First open NESSIE Workshop, Leuven, Belgium, volume 13, page 14. Citeseer.

[11] Barreto, P., Rijmen, V., et al. (2000b). The whirlpool hashing function. In First open NESSIE Workshop, Leuven, Belgium, volume 13, page 14. Citeseer.

[12] Becker, M. and Desoky, A. (2004). A study of the dvd content scrambling system (css) algorithm. In Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, 2004., pages 353–356. IEEE.

[13] Bernasconi, J. (1987). Low autocorrelation binary sequences: statistical mechanics and configuration space analysis. Journal de Physique, 48(4):559–567.

[14] Berry, D. A. and Fristedt, B. (1985). Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability). London: Chapman and Hall, 5:71–87.

[15] Bhattacharya, D., Bansal, N., Banerjee, A., and RoyChowdhury, D. (2007). A near optimal S-box design. In International Conference on Information Systems Security, pages 77–90. Springer.

[16] Biham, E. (1994). On Matsui's linear cryptanalysis. In Workshop on the Theory and Application of of Cryptographic Techniques, pages 341–355. Springer.

[17] Biham, E. and Shamir, A. (1991). Differential cryptanalysis of DES-like cryptosystems. Journal of CRYPTOLOGY, 4(1):3–72.

[18] Bikov, D., Bouyukliev, I., and Bouyuklieva, S. (2019). Bijective S-boxes of different sizes obtained from quasi-cyclic codes. Journal of Algebra Combinatorics Discrete Structures and Applications, 6(3):123–134.

[19] Biryukov, A. and Perrin, L. (2015). On reverse-engineering s-boxes with hidden design criteria or structure. In Annual Cryptology Conference, pages 116–140. Springer.

[20] Biryukov, A., Perrin, L., and Udovenko, A. (2016). Reverse-engineering the s-box of streebog, kuznyechik and stribobr1. In Annual International Conference on the Theory and Applications of Cryptographic Techniques, pages 372–402. Springer.

[21] Borwein, P., Choi, K.-K., and Jedwab, J. (2004). Binary sequences with merit factor greater than 6.34. IEEE transactions on information theory, 50(12):3234–3249.

[22] Boškovic, B., Brglez, F., and Brest, J. (2016). A github archive for solvers and solutions of the labs problem. For updates, see https://github. com/borkob/git_labs (January 2016).

[23] Bošković, B., Brglez, F., and Brest, J. (2017). Low-autocorrelation binary sequences: On improved merit factors and runtime predictions to achieve them. Applied Soft Computing, 56:262–285.

[24] Bouyukliev, I., Bikov, D., and Bouyuklieva, S. (2017). S-boxes from binary quasi-cyclic codes. Electronic Notes in Discrete Mathematics, 57:67–72.

[25] Brest, J. and Bošković, B. (2018). A heuristic algorithm for a low autocorrelation binary sequence problem with odd length and high merit factor. IEEE Access, 6:4127–4134.

[26] Brest, J. and Bošković, B. (2020). In searching of long skew-symmetric binary sequences with high merit factors. arXiv preprint arXiv:2011.00068.

[27] Byrnes, J. and Newman, D. J. (1990). The l4 norm of a polynomial with coefficients±1. Amer. Math. Monthly, 97:42–45.

[28] Canteaut, A., Duval, S., and Leurent, G. (2015a). Construction of lightweight S-boxes using Feistel and MISTY structures. In International Conference on Selected Areas in Cryptography, pages 373–393. Springer.

[29] Canteaut, A., Duval, S., and Leurent, G. (2015b). Construction of lightweight s-boxes using feistel and misty structures. In International Conference on Selected Areas in Cryptography, pages 373–393. Springer.

[30] Chen, G. (2008). A novel heuristic method for obtaining S-boxes. Chaos, Solitons & Fractals, 36(4):1028–1036.

[31] Clark, J. A., Jacob, J. L., and Stepney, S. (2005). The design of S-boxes by simulated annealing. New Generation Computing, 23(3):219–231.

[32] Coppersmith, D. (1994). The data encryption standard (des) and its strength against attacks. IBM journal of research and development, 38(3):243–250.

[33] Courtois, N. T. and Pieprzyk, J. (2002). Cryptanalysis of block ciphers with overdefined systems of equations. In International Conference on the Theory and Application of Cryptology and Information Security, pages 267–287. Springer.

[34] Coxson, G. and Russo, J. (2005). Efficient exhaustive search for optimal-peak-sidelobe binary codes. IEEE Transactions on Aerospace and Electronic Systems, 41(1):302–308.

[35] Coxson, G. E., Hill, C. R., and Russo, J. C. (2014). Adiabatic quantum computing for finding low-peak-sidelobe codes. In 2014 IEEE High Performance Extreme Computing Conference (HPEC), pages 1–6. IEEE.

[36] Coxson, G. E., Russo, J. C., and Luther, A. (2020). Long low-psl binary codes by multi-thread evolutionary search. In 2020 IEEE International Radar Conference (RADAR), pages 256–261. IEEE.

[37] Cui, L. and Cao, Y. (2007). A new S-box structure named Affine-Power-Affine. International Journal of Innovative Computing, Information and Control, 3(3):751–759.

[38] Daemen, J. and Rijmen, V. (2013). The design of Rijndael: AES-the advanced encryption standard. Springer Science & Business Media.

[39] De Groot, C., Würtz, D., and Hoffmann, K. H. (1992). Low autocorrelation binary sequences: Exact enumeration and optimization by evolutionary strategies. Optimization, 23(4):369–384.

[40] de la Cruz Jiménez, R. A. (2017). Generation of 8-Bit S-Boxes Having Almost Optimal Cryptographic Properties Using Smaller 4-Bit S-Boxes and Finite Field Multiplication. In International Conference on Cryptology and Information Security in Latin America, pages 191–206. Springer.

[41] Developers, T. S. (2016). Sagemath.

[42] Dimitrov, M. (2020a). On the aperiodic autocorrelations of rotated binary sequences. IEEE Communications Letters, 25(5):1427–1430.

[43] Dimitrov, M., Baicheva, T., and Nikolov, N. (2021). Hybrid constructions of binary sequences with low autocorrelation sideobes. IEEE Access, 9:112400–112410.

[44] Dimitrov, M., Baitcheva, T., and Nikolov, N. (2020). On the generation of long binary sequences with record-breaking PSL values. IEEE Signal Processing Letters, 27:1904–1908.

[45] Dimitrov, M. M. (2020b). On the design of chaos-based s-boxes. IEEE Access, 8:117173–117181.

[46] Dimitrov, M. M. (2021). A framework for fine-grained nonlinearity optimization of boolean and vectorial boolean functions. IEEE Access, 9:124910–124920.

[47] Dmitriev, D. and Jedwab, J. (2007). Bounds on the growth rate of the peak sidelobe level of binary sequences. Advances in Mathematics of Communications, 1(4):461.

[48] Dolmatov, V. (2016). Gost r 34.12-2015: Block cipher "kuznyechik". Transformation, 50:10.

[49] Du, K. L., Wu, W. H., and Mow, W. H. (2013). Determination of long binary sequences having low autocorrelation functions. US Patent 8,493,245.

[50] FIPS, P. (1999). 46-3. data encryption standard (des). National Institute of Standards and Technology, 25(10):1–22.

[51] Flynn, M. J. (1972). Some computer organizations and their effectiveness. IEEE transactions on computers, 100(9):948–960.

[52] Gallardo, J. E., Cotta, C., and Fernández, A. J. (2009). Finding low autocorrelation binary sequences with memetic algorithms. Applied Soft Computing, 9(4):1252–1262.

[53] Gérard, B., Grosso, V., Naya-Plasencia, M., and Standaert, F.-X. (2013). Block ciphers that are easier to mask: How far can we go? In International Workshop on Cryptographic Hardware and Embedded Systems, pages 383–399. Springer.

[54] Gilbert, H. and Peyrin, T. (2010). Super-Sbox cryptanalysis: improved attacks for AES-like permutations. In International Workshop on Fast Software Encryption, pages 365–383. Springer.

[55] Golay, M. (1972). A class of finite binary sequences with alternate auto-correlation values equal to zero (corresp.). IEEE Transactions on Information Theory, 18(3):449–450.

[56] Golay, M. (1975). Hybrid low autocorrelation sequences (corresp.). IEEE Transactions on Information Theory, 21(4):460–462.

[57] Golay, M. (1977). Sieves for low autocorrelation binary sequences. IEEE Transactions on information theory, 23(1):43–51.

[58] Golay, M. (1982). The merit factor of long low autocorrelation binary sequences (corresp.). IEEE Transactions on Information Theory, 28(3):543–549.

[59] Golay, M. (1983). The merit factor of legendre sequences (corresp.). IEEE Transactions on Information Theory, 29(6):934–936.

[60] Golay, M. J. and Harris, D. B. (1990). A new search for skewsymmetric binary sequences with optimal merit factors. IEEE Transactions on Information Theory, 36(5):1163–1166.

[61] Gold, R. (1967). Optimal binary sequences for spread spectrum multiplexing (Corresp.). IEEE Transactions on Information Theory, 13(4):619–621.

[62] Golomb, S. W. et al. (1967). Shift register sequences. Aegean Park Press.

[63] Grosso, V., Leurent, G., Standaert, F.-X., and Varıcı, K. (2014a). LS-designs: Bitslice encryption for efficient masked software implementations. In International Workshop on Fast Software Encryption, pages 18–37. Springer.

[64] Grosso, V., Leurent, G., Standaert, F.-X., Varici, K., Durvaux, F., Gaspar, L., and Kerckhof, S. (2014b). Scream & iscream side-channel resistant authenticated encryption with masking. Submission to CAESAR.

[65] Gurobi Optimization, I. (2018). Gurobi optimizer reference manual. URL http://www. gurobi. com.

[66] Halim, S., Yap, R. H., and Halim, F. (2008). Engineering stochastic local search for the low autocorrelation binary sequence problem. In International Conference on Principles and Practice of Constraint Programming, pages 640–645. Springer.

[67] He, H., Stoica, P., and Li, J. (2009). Designing unimodular sequence sets with good correlations—including an application to mimo radar. IEEE Transactions on Signal Processing, 57(11):4391–4405.

[68] Heys, H. M. (2002). A tutorial on linear and differential cryptanalysis. Cryptologia, 26(3):189–221.

[69] Hoholdt, T. and Jensen, H. E. (1988). Determination of the merit factor of legendre sequences. IEEE Transactions on Information Theory, 34(1):161–164.

[70] Isa, H., Jamil, N., and Z'aba, M. R. (2013). S-box construction from non-permutation power functions. In Proceedings of the 6th International Conference on Security of Information and Networks, pages 46–53. ACM.

[71] Isa, H., Jamil, N., and Z'aba, M. R. (2016). Construction of cryptographically strong S-Boxes inspired by bee waggle dance. New generation computing, 34(3):221–238.

[72] Ivanov, G., Nikolov, N., and Nikova, S. (2015). Cryptographically strong S-boxes generated by modified immune algorithm. In International Conference on Cryptography and Information Security in the Balkans, pages 31–42. Springer.

[73] Jakobsen, T. and Knudsen, L. R. (1997). The interpolation attack on block ciphers. In International Workshop on Fast Software Encryption, pages 28–40. Springer.

[74] Jedwab, J. (2004). A survey of the merit factor problem for binary sequences. In International Conference on Sequences and Their Applications, pages 30–55. Springer.

[75] Jedwab, J. and Yoshida, K. (2006). The peak sidelobe level of families of binary sequences. IEEE transactions on information theory, 52(5):2247–2254.

[76] Junod, P. and Vaudenay, S. (2004). Fox: a new family of block ciphers. In International Workshop on Selected Areas in Cryptography, pages 114–129. Springer.

[77] Karpman, P. and Grégoire, B. (2016). The littlun s-box and the fly block cipher. In Lightweight Cryptography Workshop, pages 17–18.

[78] Kasami, T. (1966). Weight distribution formula for some class of cyclic codes. Coordinated Science Laboratory Report no. R-285.

[79] Kazymyrov, O., Kazymyrova, V., and Oliynykov, R. (2013). A Method For Generation Of High-Nonlinear S-Boxes Based On Gradient Descent. IACR Cryptology ePrint Archive, 2013:578.

[80] Kerahroodi, M. A., Aubry, A., De Maio, A., Naghsh, M. M., and Modarres-Hashemi, M. (2017). A coordinate-descent framework to design low psl/isl sequences. IEEE Transactions on Signal Processing, 65(22):5942–5956.

[81] Leukhin, A., Parsaev, N., Bezrodnyi, V., and Kokovihina, N. (2017). The exhaustive search for optimum minimum peak sidelobe binary sequences. Bulletin of the Russian Academy of Sciences: Physics, 81(5):575–578.

[82] Leukhin, A. and Potehin, E. (2012). Binary sequences with minimum peak sidelobe level up to length 68. arXiv preprint arXiv:1212.4930.

[83] Leukhin, A. and Potekhin, E. (2015). A Bernasconi model for constructing ground-state spin systems and optimal binary sequences. In Journal of Physics: Conference Series, volume 613, page 012006. IOP Publishing.

[84] Leukhin, A. N. and Potekhin, E. N. (2013). Optimal peak sidelobe level sequences up to length 74. In 2013 European Radar Conference, pages 495–498. IEEE.

[85] Leukhin, Anatolii N and Potekhin, Egor N (2014). Exhaustive search for optimal minimum peak sidelobe binary sequences up to length 80. In International Conference on Sequences and Their Applications, pages 157–169. Springer.

[86] Levanon, N. and Mozeson, E. (2004). Radar signals. John Wiley & Sons.

[87] Lim, C. H. (1998). Crypton: A new 128-bit block cipher. NIsT AEs Proposal.

[88] Lim, C. H. (1999). A revised version of crypton: Crypton v1. 0. In International Workshop on Fast Software Encryption, pages 31–45. Springer.

[89] Lin, R., Soltanalian, M., Tang, B., and Li, J. (2019). Efficient design of binary sequences with low autocorrelation sidelobes. IEEE Transactions on Signal Processing, 67(24):6397–6410.

[90] Lindner, J. (1975). Binary sequences up to length 40 with best possible autocorrelation function. Electronics letters, 11(21):507–507.

[91] Littlewood, J. (1966). On Polynomials $\sum^n \pm z^m$, $\sum^n e^{\alpha_m i} z^m$, $z = e^{\theta i}$. Journal of the London Mathematical Society, 1(1):367–376.

[92] Losanitsch, S. (1897). Die isomerie-arten bei den homologen der paraffin-reihe. Berichte der deutschen chemischen Gesellschaft, 30(2):1917–1926.

[93] Madras, N. and Slade, G. (2013). The self-avoiding walk. Springer Science & Business Media.

[94] Mamadolimov, A., Isa, H., and Mohamad, M. S. (2013). Practical bijective S-box design. arXiv preprint arXiv:1301.4723.

[95] Massey, J. L. (1993). Safer k-64: A byte-oriented block-ciphering algorithm. In International Workshop on Fast Software Encryption, pages 1–17. Springer.

[96] Matsui, M. (1993). Linear cryptanalysis method for DES cipher. In Workshop on the Theory and Application of of Cryptographic Techniques, pages 386–397. Springer.

[97] Meier, W. and Staffelbach, O. (1989). Nonlinearity criteria for cryptographic functions. In Workshop on the Theory and Application of Cryptographic Techniques, pages 549–562. Springer.

[98] Mertens, S. (1996). Exhaustive search for low-autocorrelation binary sequences. Journal of Physics A: Mathematical and General, 29(18):L473.

[99] Militzer, B., Zamparelli, M., and Beule, D. (1998). Evolutionary search for low autocorrelated binary sequences. IEEE Transactions on Evolutionary Computation, 2(1):34–39.

[100] Millan, W. (1998). How to improve the nonlinearity of bijective S-boxes. In Australasian Conference on Information Security and Privacy, pages 181–192. Springer.

[101] Millan, W., Burnett, L., Carter, G., Clark, A., and Dawson, E. (1999). Evolutionary heuristics for finding cryptographically strong S-boxes. In International Conference on Information and Communications Security, pages 263–274. Springer.

[102] Mow, W. H., Du, K.-L., and Wu, W. H. (2015). New evolutionary search for long low autocorrelation binary sequences. IEEE Transactions on aerospace and electronic systems, 51(1):290–303.

[103] Mroczkowski, P. (2009). Generating Pseudorandom S-Boxes-a Method of Improving the Security of Cryptosystems Based on Block Ciphers. Journal of Telecommunications and Information Technology, pages 74–79.

[104] Nasrabadi, M. A. and Bastani, M. H. (2010). A survey on the design of binary pulse compression codes with low autocorrelation. In Trends in Telecommunications Technologies. IntechOpen.

[105] Nunn, C. J. and Coxson, G. E. (2008). Best-known autocorrelation peak sidelobe levels for binary codes of length 71 to 105. IEEE transactions on Aerospace and Electronic Systems, 44(1):392–395.

[106] Nyberg, K. (1991a). Perfect nonlinear S-boxes. In Workshop on the Theory and Application of Cryptographic Techniques, pages 378–386. Springer.

[107] Nyberg, K. (1991b). Perfect nonlinear S-boxes. In Workshop on the Theory and Application of of Cryptographic Techniques, pages 378–386. Springer.

[108] Oliphant, T. E. (2006). A guide to NumPy, volume 1. Trelgol Publishing USA.

[109] Oliynykov, R., Gorbenko, I., Kazymyrov, O., Ruzhentsev, V., Kuznetsov, O., Gorbenko, Y., Dyrda, O., Dolgov, V., Pushkaryov, A., Mordvinov, R., et al. (2015). A new encryption standard of ukraine: The kalyna block cipher. IACR Cryptology ePrint Archive, 2015:650.

[110] Orhanou, G., El Hajji, S., and Bentaleb, Y. (2010). Snow 3g stream cipher operation and complexity study. Contemporary Engineering Sciences-Hikari Ltd, 3(3):97–111.

[111] Packebusch, T. and Mertens, S. (2016). Low autocorrelation binary sequences. Journal of Physics A: Mathematical and Theoretical, 49(16):165001.

[112] Perrin, L. P. (2017). Cryptanalysis, reverse-engineering and design of symmetric cryptographic algorithms. PhD thesis, University of Luxembourg, Luxembourg, Luxembourg.

[113] Perrin, L. P. and Udovenko, A. (2017). Exponential s-boxes: a link between the s-boxes of belt and kuznyechik/streebog. IACR Transactions on Symmetric Cryptology, 2016(2):99–124.

[114] Picek, S., Cupic, M., and Rotim, L. (2016a). A new cost function for evolution of s-boxes. Evolutionary computation, 24(4):695–718.

[115] Picek, S., Santana, R., and Jakobovic, D. (2016b). Maximal nonlinearity in balanced boolean functions with even number of inputs, revisited. In 2016 IEEE Congress on Evolutionary Computation (CEC), pages 3222–3229. IEEE.

[116] Piret, G., Roche, T., and Carlet, C. (2012). Picaro–a block cipher allowing efficient higher-order side-channel resistance. In International Conference on Applied Cryptography and Network Security, pages 311–328. Springer.

[117] Pott, A. (2006). Finite geometry and character theory. Springer.

[118] Prestwich, S. D. (2013). Improved branch-and-bound for low autocorrelation binary sequences. arXiv preprint arXiv:1305.6187.

[119] Reeds III, J. A. (1992). Cryptosystem for cellular telephony. US Patent 5,159,634.

[120] Rijmen, V. and Barreto, P. (2000). The anubis block cipher. Submission to NESSIE.

[121] Rijmen, V. and Preneel, B. (1997). A family of trapdoor ciphers. In International Workshop on Fast Software Encryption, pages 139–148. Springer.

[122] Rudin, W. (1959). Some theorems on fourier coefficients. Proceedings of the American Mathematical Society, 10(6):855–859.

[123] Rushanan, J. J. (2006). Weil sequences: A family of binary sequences with good correlation properties. In 2006 IEEE International Symposium on Information Theory, pages 1648–1652. IEEE.

[124] SageMath. Preliminary State Standard of Republic of Belarus (STB P 34.101.31–2007). http://apmi.bsu.by/assets/files/std/belt-spec27.pdf.

[125] SageMath. SageMath Sbox library. https://github.com/sagemath/sage/blob/master/src/sage/crypto/sboxes.py.

[126] Sarkar, P. and Maitra, S. (2000). Nonlinearity bounds and constructions of resilient boolean functions. In Annual International Cryptology Conference, pages 515–532. Springer.

[127] Schneier, B., Kelsey, J., Whiting, D., Wagner, D., Hall, C., and Ferguson, N. (1998). Twofish: A 128-bit block cipher. aes submission, 15 june 1998.

[128] Schotten, H. D. and Lüke, H. D. (2005). On the search for low correlated binary sequences. AEU-International Journal of Electronics and Communications, 59(2):67–78.

[129] Shapiro, H. S. (1952). Extremal problems for polynomials and power series. PhD thesis, Massachusetts Institute of Technology.

[130] Shirai, T., Shibutani, K., Akishita, T., Moriai, S., and Iwata, T. (2007). The 128-bit blockcipher clefia. In International workshop on fast software encryption, pages 181–195. Springer.

[131] Skipjack, N. (1998). KEA algorithm specifications. Online document: http://csrc. nist. org/encryption/skipjack/skipjack. pdf.

[132] Skolnik, M. I. (1970). Radar handbook.

[133] Soltanalian, M. and Stoica, P. (2012). Computational design of sequences with good correlation properties. IEEE Transactions on Signal processing, 60(5):2180–2193.

[134] Song, J., Babu, P., and Palomar, D. P. (2015). Sequence design to minimize the weighted integrated and peak sidelobe levels. IEEE Transactions on Signal Processing, 64(8):2051–2064.

[135] Souravlias, D., Parsopoulos, K. E., and Meletiou, G. C. (2017). Designing bijective S-boxes using Algorithm Portfolios with limited time budgets. Applied Soft Computing, 59:475–486.

[136] Standaert, F.-X., Piret, G., Rouvroy, G., Quisquater, J.-J., and Legat, J.-D. (2004). Iceberg: An involutional cipher efficient for block encryption in reconfigurable hardware. In International Workshop on Fast Software Encryption, pages 279–298. Springer.

[137] Stern, J. and Vaudenay, S. (1998). Cs-cipher. In International Workshop on Fast Software Encryption, pages 189–204. Springer.

[138] Tesař, P. (2010). A new method for generating high non-linearity s-boxes. Radioengineering, 19(1):23–26.

[139] Turyn, R. et al. (1968). Sequences with small correlation. In Error correcting codes, pages 195–228. Wiley New York.

[140] Wagner, D. (1999). The boomerang attack. In International Workshop on Fast Software Encryption, pages 156–170. Springer.

[141] Watanabe, D., Furuya, S., Yoshida, H., Takaragi, K., and Preneel, B. (2002). A new keystream generator mugi. In International Workshop on Fast Software Encryption, pages 179–194. Springer.

[142] Wiener, N. (1964). Extrapolation, interpolation, and smoothing of stationary time series. The MIT press.

[143] Wikipedia source (1999). Wikipedia. https://en.wikipedia.org/wiki/Iraqi_block_cipher.

[144] Wu, H., Bao, F., Deng, R. H., and Ye, Q.-Z. (1998). Cryptanalysis of rijmen-preneel trapdoor ciphers. In International Conference on the Theory and Application of Cryptology and Information Security, pages 126–132. Springer.

[145] Xiu-tao, F. (2011). Zuc algorithm: 3gpp lte international encryption standard [j]. Information Security and Communications Privacy, 12.