Логический криптоанализ в исследовании стойкости одной схемы кодовой подписи

И. В. Чижов, Н. П. Ташевцева

Аннотация—**В работе предлагается алгоритм, который по входу задачи криптоаналитика поиска секретных ключей схемы электронно-цифровой подписи pqsigRM строит эквивалентный ему вход задачи ВЫПОЛНИМОСТИ КНФ. Доказано, что этот алгоритм является полиномиальным. Приводятся теоретические оценки параметров полученной КНФ – длины и количества входящих в нее переменных. В ходе исследования была предложена практическая реализация на Python алгоритма сводимости, то есть алгоритм создания соответствующей КНФ в DIMACS формате для любых параметров r, m схемы pqsigRM. В работе приводятся результаты экспериментов запуска этой реализации на некоторых параметрах схемы и результаты экспериментов по решению задачи выполнимости полученных КНФ для некоторых параметров r, m изначальной задачи криптоаналитика, различными SAT-решателями с открытым исходным кодом - победителями и призерами конкурса SAT Competition 2018, SAT Race 2019, а также более ранними версиями, которые были использованы ранее для криптоанализа системы Мак-Элиса. Описаны параметры схемы ЭЦП pqsigRM, для которых атака применима, т.е. можно подобрать секретные ключи.**

Ключевые слова—**Постквантовая криптография, полиномиальная атака, электронно-цифровая подпись, коды Рида-Маллера**

I. Введение

Стойкость всех классических криптосистем, получивших широкое распространение на практике, основана на сложности решения некоторых теоретико-числовых задач. Примером могут служить задачи - «претенденты» на односторонние функции: задачи факторизации числа или задачи дискретного логарифмирования в конечной группе. Однако, еще в 1994 году был разработан квантовый алгоритм Шора [1], позволяющий восстановить аргументы односторонних функций за приемлемое время, тем самым давая возможность взламывать некоторые классические криптосистемы.

С каждым годом мы становимся все ближе к созданию квантового компьютера [2], [3]. Помимо того, что существует множество прототипов, еще в 2000-х были созданы первые небольшие регистры [4], состоящие всего из нескольких квантовых битов, и первые 5-битные квантовые вычислительные системы. Пока непонятно, когда именно будут созданы полноценные квантовые вычислительные системы, однако, некоторые ученые, работающие над их разработкой, предсказывают создание квантового компьютера, способного взломать 2000-битный RSA, до 2030 года [5]. В случае появления многокубитного квантового компьютера все современные криптографические механизмы с открытым ключом станут нестойкими.

Национальный институт стандартов и технологий США (NIST) уже официально объявил о необходимости перехода к криптографии, устойчивой к атакам, использующим квантовый компьютер. В России Техническим комитетом 26 по стандартизации «Криптографическая защита информации» создана отдельная группа по «Постквантовым криптографическим механизмам», занимающаяся исследованиями и вопросами стандартизации в области постквантовой криптографии.

Также вопрос о переходе к исследованию альтернативных к существующим криптосистемам, подогревается т ем, что особое распространение в последние годы получила технология «Блокчейн», для работы которой используются схемы коллективной подписи на основе ассиметричного шифрования и протоколы обмена ключами. При этом надежность технологии полностью зависит от стойкости криптографических механизмов, основывающейся на сложности решения задач факторизации чисел и дискретного логарифмирования. Тогда при появлении многокубитного квантового компьютера использование всех существующих сейчас в мире криптографических валют станет ненадежным. Поэтому все чаще поднимается вопрос об исследовании и переходе к использованию более надежных постквантовых криптографических механизмов.

Одними из основных претендентов на замену являются кодовые криптосистемы Мак-Элиса и Нидеррайтера, стойкость которых основана на сложности задачи декодирования полных линейных кодов, а в качестве проверочной (или порождающей) матрицы берется некоторая случайная матрица. В 1978 году Роберт Мак-Элис предложил первую криптосистему, основанную на алгебраическом блоковом кодировании и использующую в процессе шифрования рандомизацию [6]. Эта криптосистема использует двоичные коды Гоппы, и идея ее построения основана на максировке линейного кода с эффективным алгоритмом декодирования под случайный код без видимой структуры, а как известно, общая задача декодирования являестя NP-полной [7]. Г. Нидеррайтер в 1986 г. предложил модифицировать криптосистему Мак-Элиса, заменив коды Гоппы на коды Рида-Соломона [8]. Плюсом системы Нидеррайтера является то, что она может быть использована для создания ЭЦП. Однако такая система, построенная на кодах Рида-Соломона, оказалась нестойкой и в 1992 году В. М. Сидельников и С. О. Шестаков построили атаку [9] – показали, что можно восстановить структуру секретного ключа по открытому

Статья получена 25 сентября 2020

Иван Владимирович Чижов, МГУ им. М. В. Ломоносова, Федеральный исследовательский центр «Информатика и управление» РАН, АО «НПК «Криптонит»» (email: ichizhov@cs.msu.ru).

Наталия Павловна Ташевцева, МГУ им. М.В. Ломоносова, ООО «АТ Груп», (email: ntashevtseva@gmail.com).

Работа частично поддержана грантом РФФИ №182903124мк

и подобрать такие новые секретные матрицы S', H', что $H_{pub} = S'H'$. Также В. М. Сидельниковым в 1994 году было показано, что атака на систему Нидеррайтера может быть полиномиально сведена к атаке на систему Мак-Элиса и обратно [10].

Поэтому В. М. Сидельников в 1994 году предложил заменить коды Гоппы в системе Мак-Элиса на коды Рида-Маллера. Однако, криптографический анализ [10] этой системы, а позднее субэкспоненциальный алгоритм атаки Миндера-Шокролахи [11] и полиномальный алгоритм атаки Чижова-Бородина [12] для большого числа практических параметров, доказали ее недостаточную стойкость.

В 2016 году на конкурсе квантово-устойчивых криптографических механизмов, организованном NIST с целью выбора американского стандарта, была представлена новая ЭЦП pqsigRM, основанная на модифицированных кодах Рида-Маллера [13]. Эта схема является улучшением теоретико-кодовой схемы подписи, разработанной Н. Куртуа, М. Финиазом и Н. Сендриером (N. Courtois, M. Finiazs, N. Sendrier - CFS) [14]. CFS является первым стойким алгоритмом формирования и проверки ЭЦП, использующим алгебраические коды, до этого считалось, что такую схему, например, на основе криптосистемы Мак-Элиса, построить невозможно. CFS протокол ЭЦП основан на криптосистеме Нидеррайтера и его стойкость сводится к сложности решения задачи синдромного декодирования и задачи различимости перестановочных кодов Гоппы от случайных. ЭЦП по схеме CFS имеет высокую стойкость к атакам с использованием квантового вычислителя, в то же время ее очевидным недостатком является высокая трудоемкость формирования ЭЦП. Для формирования подписи по схеме CFS необходимо найти синдром, соответствующий вектору ошибки веса, не превосходящего исправляющей способности кода, поэтому трудоёмкость формирования ЭЦП существенно зависит от вероятности появления такого вектора в случайной выборке из всего пространства и растет как факториал от исправляющей способности кода. Коды Рида-Маллера исправляют много ошибок, поэтому их использование значительно бы улучшило скорость формирования ЭЦП.

В работе [13] описывается модификация схемы подписи, где для начала предлагается заменить коды Гоппы на коды Рида-Маллера в CFS. Однако только замена кодов Гоппы РМ-кодами в CFS не гарантирует устойчивость к некоторым известным атакам. Криптосистема Мак-Элиса, основанная на кодах Рида-Маллера, не устойчива к атаке Миндера-Шокролахи и атаке Чижова-Бородина. Поэтому для защиты от этих атак предлагается модификация, главная идея которой заключается в формировании матрицы кода, на основе которого строится схема подписи, из порождающих матриц кода Рида-Маллера, испорченных частичной перестановкой их столбцов.

В работе проводится логический криптоанализ схемы ЭЦП pqsigRM. Основная его идея состоит в сведении решения задачи криптоаналитика к поиску вектора, на котором выполнима БФ, представленная в виде КНФ - важному виду задачи теории сложности «о ВЫПОЛНИМОСТИ БУЛЕВОЙ ФОРМУЛЫ» (SAT задача - «SATisfiability problem»). Несмотря на то, что С. Куком в 1971 г. была доказана NP-полнота этой задачи [15], существуют подходы, позволяющие для некоторых типов КНФ ускорять алгоритмы определения выполнимости БФ, а также использовать распределенные алгоритмы, что дает возможность параллелизовать исходную задачу криптоаналитика. Именно эти подходы используются для создания SAT-решателей - программ, позволяющих найти решение задачи ВЫПОЛНИМОСТИ КНФ. История их развития началась еще до доказательства теоремы Кука-Левина, в 1960-х годах с создания Дэвисом и Путнамом классических дедуктивных методов для решения SAT задачи и продолжается до сих пор. Каждый год проводится конкурс SAT-решателей в различных номинациях.

Возжность применения логического криптоанализа к кодовым системам исследовалась ранее в работе М. Бородина [16]. В ней была предложена атака на криптосистему Мак-Элиса, построенную на основе двоичных кодов Рида-Маллера, посредством сведения задачи нахождения секретного ключа к задаче ВЫПОЛНИМОСТИ КНФ, и установлено, что параметры получающейся КНФ не превосходят полинома 2-й степени от длины кода.

В настоящей работе предлагается использовать особенности структуры порождающей матрицы новой схемы ЭЦП pqsigRM и особенности процесса генерации публичного ключа для сведения исходной задачи криптоаналитика поиска ключей формирования подписи к задаче ВЫПОЛНИМОСТЬ КНФ, строится полиномиальный алгоритм, который по входу задачи криптоаналитика строит эквивалентный ему вход задачи ВЫПОЛНИМОСТИ КНФ. Приводятся результаты экспериментов запуска этой реализации на некоторых параметрах. Помимо этого в работе представлены результаты экспериментов по решению задачи выполнимости полученных КНФ для некоторых параметров изначальной задачи криптоаналитика, различными open-source SAT-решателями – победителями и призерами конкурса SAT Competition 2018, SAT Race 2019, а также более ранними версиями, которые были использованы в работе [16] для криптоанализа системы Мак-Элиса. Также установлены некоторые параметры схемы ЭЦП pqsigRM, для которых атака применима, то есть можно подобрать секретные ключи.

II. Коды Рида-Маллера

Пусть $x = (x_1, ..., x_n)$ и $y = (y_1, ..., y_n)$ — векторы над полем F_2. Тогда обозначим скалярное произведение векторов x и y как число $x \cdot y \in F_2 : x \cdot y = x_1 \cdot y_1 \oplus x_2 \cdot y_2 + ... \oplus x_n \cdot y_n$, где $\oplus$ — сумма по по модулю 2. Векторы, скалярное произведение которых равно нулю, принято называть ортогональными.

Определение 1. Линейный $[n, k]$ – код является произвольным линейным подпространством C размерности k пространства F_q^n, где F_q – конечное поле из q элементов. Параметр n — длина кода, k — размерность кода.

Рассмотрим матрицу $G \in F_q^{k \times n}$, состоящую из k линейно независимых строк. Тогда линейная оболочка C строк матрицы G является линейным $[n, k]$ – кодом, элементы $x \in C$ — *кодовыми словами*, матрица G — *порождающей матрицей* $[n, k]$ – кода.

Определение 2. Систематической (канонической) формой порождающей матрицы линейного кода будем называть матрицу вида $\hat{G} = [I_k \mid A]$, получающуюся из

порождающей матрицы G проведением операций только над строками (или только над столбцами).

Определение 3. [17] Если L – линейный $[n,k]$ - код, то *дуальный (ортогональный)* к нему код $L^{\perp}$ определяется как множество всех векторов, ортогональных всем кодовым словам кода L:

$$L^{\perp} = \{u \mid u \cdot v = 0 \ \forall v \in L\}$$

Определение 4. [17] *Синдромом вектора* y называется вектор $s = Hy^T$, где H - проверочная матрица, $y \in V_n, H \in V^{(n-k)\times n}, s \in V_{n-k}$.

Расстояние Хэмминга между двумя векторами – это количество позиций, в которых они отличаются.

Вес Хэмминга вектора – это количество ненулевых позиций этого вектора. Обозначается $\mathrm{wt}(x)$.

Кодовым расстоянием называется

$$d_{min} = \min\{\mathrm{dist}(u,v)\} = \min\{\mathrm{wt}(u-v)\}, u \neq v$$

Теорема II.1. *[17] Пусть C_1 и C_2 двоичные $[n,k_1]$ и $[n,k_2]-$ коды с кодовыми расстояниями d_1 и d_2 соответственно. Тогда множество*

$$C = \{(x|x \oplus y) \ : \ x \in C_1, y \in C_2\}$$

является $[n, k_1 + k_2]-$кодом с кодовым расстоянием $\min\{d_1, d_2\}$.

Синдромом вектора y называется вектор $s = Hy^T$, где H – проверочная матрица, $y \in V_n, H \in V^{(n-k)\times n}, s \in V_{n-k}$.

Определение 5. Кодом Рида-Маллера $RM(r,m)$ для любого $r : 0 \leq r \leq m$, называется множество векторзначений Ω_f всех булевых функций $f(y_1, ..., y_m)$, степень нелинейности (т. е. максимальная длина монома, входящего в полином Жегалкина функции f) которых не превосходит r.

Базисом кода являются все мономы степени не большей r от m переменных:

$$1, x_1, ..., x_m, x_1 \ \& \ x_2, ..., x_{m-1} \ \& \ x_m, ..., \\ x_1 \ \& \ x_2 \ \& \ ... \ \& \ x_r, ..., x_{m-r+1} \ \& \ ... \ \& \ x_m \quad (1)$$

Определение 6. Стандартной формой порождающей матрицы кода $RM(r,m)$ будем называть матрицу, составленную из всех векторов значений мономов (1), стоящих в порядке возрастания длины мономов.

Определим далее более интутивно понятное построение порождающей матрицы кода Рида-Маллера. Пусть $G'(0,m)$ - *порождающая матрица кода* $RM(0,m)$, это строка из 2^m единиц: $G'(0,m) = (1,1,...,1)$. Пусть $\tilde{G}(1,m)$- *порождающая матрица кода* $RM(1,m)$ такая, что:

- Первая строка это матрица $G'(0,m) = (1,1,...,1)$, состоящая из 2^m единиц
- Следующие m строк и 2^m столбцов представляют из себя матрицу, в которой строки являюстя векторзначениями мономов $x_m, \ x_{m-1}, \ \ldots, x_1$, а столбец j (при нумерации от 0) соответствует двоичному представлению числа j, записанному в m битах. Обозначим эту матрицу как $G'(1,m)$.

Например, для $m = 3$ матрица будет выглядеть следующим образом:

$$G'(1,3) = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Теперь определим порождающую матрицу $\tilde{G}(r,m)$ кода $RM(r,m)$ как матрицу, составленную сверху вниз из матриц $G'(0,m), G'(1,m), ..., G'(r,m)$. При этом строками матрицы $G'(i,m)$ являются всевозможные покомпонентные произведения по i строк из $G'(1,m)$, упорядоченные лексиграфически как двоичные векторы. Число строк матрицы $G'(i,m)$ равно $\binom{m}{i}$.

Код Рида-Маллера $RM(r,m)$ является линейным $[n,k]$-кодом и состоит из 2^k слов. Его длина $n = 2^m$. Размерность $k = \sum_{i=0}^{r} \binom{m}{i}, 0 \leq r \leq m$. При $r = m$ код считается тривиальным, так как в этом случае код – это все векторы линейного пространства V_n, $n = 2^m$

Любой код Рида-Маллера $RM(r,m)$ может быть описан с помощью конструкции Плоткина следующим образом:

Теорема II.2 ([17], стр.363)**.** *Для любого кода Рида-Маллера RM(r,m)* $0 \leq r \leq m$ *верно*

$$RM(r,m) = \{(u|u \oplus v) : \\ u \in RM(r,m-1), v \in RM(r-1,m-1)\} \quad (2)$$

Теорема II.3 ([17], стр.364)**.** *Кодовое (минимальное) расстояние кода $RM(r,m)$ равно $d_{min} = 2^{m-r}$*

Теорема II.4 ([17], стр.20)**.** *Код с минимальным расстоянием d может исправлять $\lfloor \frac{d-1}{2} \rfloor$ ошибок.*

Таким образом, очевидно, что код Рида-Маллера $RM(r,m)$ может исправлять $t = 2^{m-r-1} - 1$ ошибку.

Теорема II.5 ([17], стр.365)**.** *Для всех r, $0 \leq r \leq m-1$ код $R(m-r-1,m)$ дуален коду $R(r,m)$*

III. Схема ЭЦП pqsigRM

A. Модификация кода

Благодаря тому, что код Рида-Маллера может быть определен в виде (2), его порождающую матрицу можно представить в следующей форме, два раза воспользовавшись его рекурсивной структурой:

$$G(r,m) = \begin{bmatrix} G(r,m-1) & G(r,m-1) \\ 0 & G(r-1,m-1) \end{bmatrix} =$$

$$= \left[\begin{array}{cc|cc} G(r,m-2) & G(r,m-2) & G(r,m-2) & G(r,m-2) \\ 0 & G(r-1,m-2) & 0 & G(r-1,m-2) \\ \hline 0 & 0 & G(r-1,m-2) & G(r-1,m-2) \\ 0 & 0 & 0 & G(r-2,m-2) \end{array}\right] \quad (3)$$

Определение 7. Подстановкой на множестве Ω называется любое взаимно-однозначное отображение множества Ω в себя. Множество всех подстановок на множестве Ω обозначают S_Ω. Если $\Omega = N$, $N = \{1, 2, \ldots, n\}$ то множество подстановок на N обозначается S_n

Определение 8. *Перестановочной матрицей P* называется квадратная бинарная матрица, в каждой строке

и столбце которой находится ровно один единичный элемент.

Каждая перестановочная матрица P размера $n \times n$ является матричным представлением подстановки, действующей на множестве из n элементов, $\sigma \in S_n$: элемент $P^i_j = 1 \iff \pi(i) = j$.

Определим σ_1 и σ_2 как две независимые случайные подстановки p столбцов, $p \leq \frac{n}{4}$, где n длина кода $RM(r,m)$. В схеме pqsigRM предлагается применить эти подстановки к порождающей матрице $G(r,m)$ (3) следующим образом:

- использовать первую подстановку для изменения подматриц $G(r, m-2)$, соответствующих коду $RM(r, m-2)$ во всей «первой линии» блочного представления матрицы $G(r,m)$
- вторую использовать для перестановки столбцов в подматрице $G(r-2, m-2)$, соответствующей коду $RM(r-2, m-2)$.

В итоге получаем модифицированную матрицу. Обозначим ее как $G_{\sigma_1,\sigma_2}(r,m)$:

$$\left[\begin{array}{c|c|c|c} G(r,m-2)^{\sigma_1} & G(r,m-2)^{\sigma_1} & G(r,m-2)^{\sigma_1} & G(r,m-2)^{\sigma_1} \\ \hline 0 & G(r-1,m-2) & 0 & G(r-1,m-2) \\ \hline 0 & 0 & G(r-1,m-2) & G(r-1,m-2) \\ \hline 0 & 0 & 0 & G(r-2,m-2)^{\sigma_2} \end{array}\right] \quad (4)$$

Определение 9. Кодом $RM_{\sigma_1,\sigma_2}(r,m)$ будем называть модифицированный код Рида-Маллера с порождающей матрицей $G_{\sigma_1,\sigma_2}(r,m)$ (4), получающейся из порождающей матрицы кода Рида-Маллера с помощью применения подстановок σ_1, σ_2.

B. Генерация открытого и секретного ключей

По порождающей матрице (4) схемы $RM_{\sigma_1,\sigma_2}(r,m)$ строим проверочную матрицу H_{σ_1,σ_2} следующим образом: приводим матрицу $G_{\sigma_1,\sigma_2}(r,m)$ к систематической форме с помощью метода Гаусса (если матрица такого вида существует), получаем

$$G'_{\sigma_1,\sigma_2} = [I_k \mid P].$$

Далее составляем новую матрицу, транспонированием матрицы P:

$$H_{\sigma_1,\sigma_2} = [P^T \mid I_{n-k}],$$

которая в совокупности с единичной матрицей размера $(n-k)\times(n-k)$ является искомой проверочной матрицей H_{σ_1,σ_2}.

Для генерации публичного ключа также рассмотрим матрицы:
S – невырожденная ($|A| \neq 0$), размера $(n-k)\times(n-k)$
Q – случайная перестановочная, размера $n \times n$.

Таким образом в новой схеме $pqsigRM$ будем использовать:
Открытый ключ: матрица $H_{pub} = SH_{\sigma_1,\sigma_2}Q$,
Секретные ключи: матрицы S – невырожденная, Q – перестановочная, H_{σ_1,σ_2} – проверочная для кода $RM_{\sigma_1,\sigma_2}(r,m)$ и подстановки (используемые для перестановки столбцов) σ_1 и σ_2, которые выбираются случайно и независимо друг от друга из множества подстановок специального вида.

C. Формирование и проверка подписи

Определение 10. *Криптографическая хэш-функция* $h(\cdot)$ – односторонняя функция, отображающая сообщение M произвольной длины в значение h фиксированной длины, и обладающая свойствами: устойчивости к коллизиям, устойчивости к поиску первого прообраза и устойчивости к поиску второго прообраза. Значения функции $h(\cdot)$ на словах, отличающихся друг от друга хотя бы в одном знаке, дают значительно различающиеся хэш-значения.

Выберем одну из криптографических хэш-функций

$$h(\cdot) : \{0,1\}^* \to \{0,1\}^{n-k}$$

1) Подпись сообщения: Пусть нам дано сообщение M, которое необходимо подписать. Выбираем для него случайный вектор $count \in \{0,1\}^{n-k}$.

Вычислим синдром s по следующей формуле:

$$s = h(h(M \mid H_{pub}) \mid count)$$

Нашей целью для создания подписи является найти вектор ошибки e такой, что

$$H_{pub}e^T = SH_{\sigma_1,\sigma_2}Qe^T = s$$

Домножим обе части равенства слева на матрицу, обратную к S:

$$H_{\sigma_1,\sigma_2}Q\ e^T = S^{-1}s$$

Обозначив $s' = S^{-1}s$ и $e'^T = Qe^T$, получаем:

$$H_{\sigma_1,\sigma_2}e'^T = s' \quad (5)$$

Таким образом, мы свели задачу нахождения вектора e к декодированию вектора s' (5).
Так как вместо матрицы (3) теперь используется порождающая матрица (4), то классический декодер кода Рида-Маллера (алгоритм декодирования Рида [18]) к новому коду $RM_{\sigma_1,\sigma_2}(r,m)$ не применим. Поэтому авторы новой схемы ЭЦП предлагают использовать новый алгоритм декодирования [13]. Применив его, мы находим вектор e'. Далее выполняем проверку:

$$wt(e') \leq w \quad (6)$$

1) Если соотношение (6) выполняется, то вычисляем $e^T = Q^{-1}e'^T$. Тогда подписью будет являться тройка (Сообщение, вектор ошибки, случайный вектор) $= (M, e, count)$.
2) Иначе процесс подписи признаем закончившимся неудачно и запускаем заново с выбора нового случайного вектора $count$.

2) Проверка подписи: Дана тройка $(M, e, count)$. Делаем проверку – если $wt(e') \leq w$ и $H_{pub}e^T = h(h(M \mid H_{pub}))$, то подпись принимаем, иначе считаем недействительной.

IV. Описание метода атаки и теоретические оценки

Предлагается провести атаку на схему ЭЦП с помощью техники сведения задачи нахождения секретного ключа к задаче ВЫПОЛНИМОСТИ КНФ. Для начала тонко сформулируем базовую задачу поиска секретных ключей схемы ЭЦП.

A. Базовая задача криптоаналитика

Вход: Открытый ключ H_{pub}

Задача: По данному входу получить новые секретные ключи $Q', \sigma_1', \sigma_2', S'$ такие, что

$$S' H'_{\sigma_1',\sigma_2'} Q' = H_{pub} = S H_{\sigma_1,\sigma_2} Q, \tag{7}$$

где $H'_{\sigma_1',\sigma_2'}$ – проверочная матрица для кода $RM_{\sigma_1,\sigma_2}(r,m)$.

B. Этапы сведения базовой задачи криптоаналитика поиска секретного ключа по открыму к задаче выполнимости КНФ

Процесс сведения базовой задачи криптоаналитика к задаче выполнимости КНФ предлагается разбить на несколько этапов. Обозначим основные задачи, рассматриваемые на этих этапах:

1) **Базовая задача криптоаналитика**
2) **Матричная задача криптоаналитика с соотношением**
3) **Задача решения системы булевых функций**
4) **Задача выполнимости КНФ**

Теперь подробно рассмотрим все этапы и переходы между ними.

C. Матричная задача криптоаналитика: $1 \to 2$

Переформулируем изначально поставленную задачу (7) нахождения секретных перестановок и невырожденной матрицы по открытому ключу:

$$S' H_{\sigma_1,\sigma_2} Q' = H_{pub} \iff S' H_{\sigma_1,\sigma_2} = H_{pub} Q'^{-1}$$

Переобозначим $Q = Q'^{-1}$.

Пусть G'_{σ_1,σ_2} - порождающая матрица, соответствующая порождающей матрице H_{σ_1,σ_2}. Тогда рассмотрим следующую цепочку соотношений:

$$G'_{\sigma_1,\sigma_2}(H_{pub} \cdot Q)^T = G'_{\sigma_1,\sigma_2}(S' \cdot H_{\sigma_1,\sigma_2})^T = \\ = (G'_{\sigma_1,\sigma_2} \cdot H^T_{\sigma_1,\sigma_2}) S'^T = 0$$

Таким образом, получаем систему:

$$G'_{\sigma_1,\sigma_2}(H_{pub} \cdot Q')^T = 0 \iff \\ \iff (G'_{\sigma_1,\sigma_2} \cdot Q^T) \cdot H^T_{pub} = 0,$$

эквивалентную условию (7) базовой задачи криптоаналитика.

H_{pub} известна, поэтому мы можем найти перестановки σ_1, σ_2 и перестановочную матрицу Q. Зная это, мы сможем построить проверочную H_{σ_1,σ_2}.

Тогда в уравнении $H_{pub} Q' = S' H_{\sigma_1,\sigma_2}$ будет неизвестна только матрица S', и относительно S' это выражение будет системой линейных алгебраических уравнений.

Таким образом получается, что базовая задача эквивалентна следующей матричной задаче криптоаналитика:

Вход: Открытый ключ H_{pub}

Задача: По данному входу получить новые секретные ключи Q', σ_1', σ_2' такие, что верно матричное уравнение:

$$(G'_{\sigma_1,\sigma_2} \cdot Q^T) \cdot H^T_{pub} = 0, \tag{8}$$

где матрица G'_{σ_1,σ_2} – порождающая матрица кода $RM_{\sigma_1,\sigma_2}(r,m)$, полученная из порождающей матрицы кода Рида-Маллера $RM(r,m)$ $G(r,m)$ 4 с помощью подстановок σ_1, σ_2.

D. Сведение к системе БФ: $2 \to 3$

На этом этапе сведения мы хотим получить систему булевых функций, эквивалентную матричной задаче.

Рассмотрим проверочное соотношение, полученное на предыдущем этапе (8):

$$(G'_{\sigma_1,\sigma_2} \cdot Q^T) \cdot H^T_{pub} = 0.$$

По определению порождающую матрицу кода $RM(r,m)$ можно представить в следующем виде:

$$G = \begin{pmatrix} G_0 \\ G_1 \\ \vdots \\ G_r \end{pmatrix}, \tag{9}$$

где

$$G_0 = (1,1,...,1), G_1 = \begin{pmatrix} \Omega_{y_m} \\ \vdots \\ \Omega_{y_2} \\ \Omega_{y_1} \end{pmatrix}, ..., G_r = \begin{pmatrix} \Omega_{y_{m-r} \dots y_m} \\ \vdots \\ \Omega_{y_1 y_2 \dots y_{r+1}} \\ \Omega_{y_1 y_2 \dots y_r} \end{pmatrix} \tag{10}$$

и Ω_f - это вектор-значение булевой функции f.

Рассмотрим подстановку σ_1 на множестве $\{0,...,n/4\} = \{0,...,2^{m-4}\}$, которая задаётся набором чисел.

$$\begin{pmatrix} 0 & 1 & \dots & n/4-2 & n/4-1 \\ a_0 & a_1 & \dots & a_{n/4-2} & a_{n/4-1} \end{pmatrix} = \\ = \begin{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} & \cdots & \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ \begin{pmatrix} a_0^1 \\ a_0^2 \\ \vdots \\ a_0^{m-4} \end{pmatrix} & \begin{pmatrix} a_1^1 \\ a_1^2 \\ \vdots \\ a_1^{m-4} \end{pmatrix} & \cdots & \begin{pmatrix} a_{n/4-2}^1 \\ a_{n/4-2}^2 \\ \vdots \\ a_{n/4-2}^{m-4} \end{pmatrix} & \begin{pmatrix} a_{n/4-1}^1 \\ a_{n/4-1}^2 \\ \vdots \\ a_{n/4-1}^{m-4} \end{pmatrix} \end{pmatrix}$$

Представив каждое число в двоичной записи, получим матрицу размера $m - 4 \times 2^{m-4}$. Каждую её строку можно трактовать как вектор значений некоторой булевой функции f_{m+i-3}. Поэтому очевидно, что подстановка σ^1 индуцирует отображение $\Omega_{y_i} \to \Omega_{f_i}, i = 1,...,m$.

Значит при применении подстановки σ_1 подматрица $G(r, m-2)$ переходит в матрицу такого же вида (9)-(10):

$$G(r, m-2)^{\sigma_1} = \begin{pmatrix} D_0 \\ D_1 \\ \vdots \\ D_r \end{pmatrix}, \tag{11}$$

где

$$D_0 = (1,1,...,1), D_1 = \begin{pmatrix} \Omega_{f_{m-2}} \\ \vdots \\ \Omega_{f_2} \\ \Omega_{f_1} \end{pmatrix}, ..., \\ D_r = \begin{pmatrix} \Omega_{f_{m-r-2} \dots f_{m-2}} \\ \vdots \\ \Omega_{f_1 f_2 \dots f_{r+1}} \\ \Omega_{f_1 f_2 \dots f_r} \end{pmatrix}. \tag{12}$$

Таким образом, можно представить подматрицы $G(r, m-2)^{\sigma_1}$ порождающей матрицы G'_{σ_1,σ_2} в следующем виде (13), выразив все элементы с помощью всего $(m-2)\cdot 2^{m-2}$ неизвестных, вместо 2^{2m-4} элементов при прямой замене всех элементов на неизвестные. Учитывая, что в порождающей матрице в первом ряду блоков находятся 4 одинаковые подматрицы такого типа (11)-(12), то используется всего $(m-2)\cdot 2^{m-2}$ неизвестных, вместо $2^m \cdot \sum_{i=0}^{r} C_{m-2}^i$.

$$G(r,m-2)^{\sigma_1} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ a_1^1 & a_2^1 & \dots & a_{2^{m-2}}^1 \\ \vdots & \vdots & & \vdots \\ a_1^{m-2} & a_2^{m-2} & \dots & a_{2^{m-2}}^{m-2} \\ a_1^1 a_1^2 & a_2^1 a_2^2 & \dots & a_{2^{m-2}}^1 a_{2^{m-2}}^2 \\ \vdots & \vdots & & \vdots \\ a_1^1 \dots a_1^r & a_2^1 \dots a_2^r & \dots & a_{2^{m-2}}^1 \dots a_{2^{m-2}}^r \end{pmatrix} \quad (13)$$

где вектор $(a_1^1 a_2^1 \dots a_{2^{m-2}}^1)$, например, соответствует $\Omega_{f_{m-2}}$.

Аналогичным образом можно представить и подматрицу

$$G(r-2,m-2)^{\sigma_2} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ b_1^1 & b_2^1 & \dots & b_{2^{m-2}}^1 \\ \vdots & \vdots & & \vdots \\ b_1^{m-2} & b_2^{m-2} & \dots & b_{2^{m-2}}^{m-2} \\ b_1^1 b_1^2 & b_2^1 b_2^2 & \dots & b_{2^{m-2}}^1 b_{2^{m-2}}^2 \\ \vdots & \vdots & & \vdots \\ b_1^1 \dots b_1^{r-2} & b_2^1 \dots b_2^{r-2} & \dots & b_{2^{m-2}}^1 \dots b_{2^{m-2}}^{r-2} \end{pmatrix} \quad (14)$$

Все остальные подматрицы матрицы порождающей матрицы G'_{σ_1,σ_2} известны.

Матрицу Q размера $n \times n$ представим в виде:

$$Q = \begin{pmatrix} q_1^1 & q_2^1 & \dots & q_{2^m}^1 \\ q_1^2 & q_2^2 & \dots & q_{2^m}^2 \\ \vdots & \vdots & & \vdots \\ q_1^{2^m} & q_2^{2^m} & \dots & q_{2^m}^{2^m} \end{pmatrix} \quad (15)$$

Тогда соотношение

$$(G'_{\sigma_1,\sigma_2} \cdot Q^T) \cdot H_{pub}^T = 0.$$

задает систему из $k \times (n-k)$ уравнений. При этом максимально возможная степень встречающихся в них мономов равна $r+1$ (следует из построения), число слагаемых уравнений не превышает $2^m 2^m = 2^{2m} = n^2$, а количество неизвестных равно $n(m-2)/2+n^2$. Функции представлены уравнениями вида сумм по модулю 2 мономов различных степеней (не более r+1) и константами 0,1. Сами мономы являются произведениями неболее чем r переменных a_j^i , т. е. так же, как и в определенном выше виде порождающей подматрицы (13)-(14), и переменных q_j^i.

Тем самым мы получили следующее утверждение:

Утверждение 1. *Представлен полиномиальный алгоритм, который по задаче криптоаналитика поиска секретных ключей Q, σ_1, σ_2 строит эквивалентную нелинейную систему булевых функций, длина которых не превышает n^2, количество неизвестных равно $n(m-2)/2+n^2$, а максимально возможная степень встречающихся в них мономов равна $r+1$.*

E. Сведение к задаче Выполнимости КНФ: $3 \to 4$

Этот шаг сведения предлагается разделить на два этапа:

1) **Сведение полиномиальной системы к линейной**
2) **Сведение линейной системы к КНФ**

1) I этап: Сведение полиномиальной системы к линейной: Известно, что система любой степени может быть сведена к системе степени ≤ 2 [19]. Этого можно добиться многократно применяя следующую замену:

$$\{m = wxyz\} \Rightarrow \{b = wx, c = yz, m = bc\}$$

Основная идея этого шага заключается в том, что логическое выражение

$$(w \vee \neg a)(x \vee \neg a)(y \vee \neg a)(z \vee \neg a)(a \vee \neg w \vee \neg x \vee \neg y \vee \neg z)$$

эквивалентно $a \iff (w \wedge x \wedge y \wedge z)$, которое эквивалентно уравнению над $GF(2)$

$$m = wxyz.$$

Поэтому для каждого монома системы степени > 1 мы можем ввести новую переменную, заменив моном на нее. При этом при каждой такой замене потребуется $d+1$ элементарных дизъюнктов (ЭД) длины $3d+1$.

Утверждение 2. *Количество новых переменных, необходимых для замены всех мономов степени $d > 1$ в системе нелинейных уравнений при сведении этой системы к линейной, не превышает:*

$$N_0 = \begin{cases} n^2 \cdot (m-2) & \text{для } r = 1 \\ n^2 \cdot \left[\frac{5}{4}\sum_{i=1}^{r-2}\binom{m-2}{i-1} + \binom{m-2}{r-1} + \binom{m-2}{r}\right] & \text{для } r > 1 \end{cases} \quad (16)$$

Доказательство. Рассмотрим построение новой матрицы $G'_{\sigma_1,\sigma_2} \cdot Q$. При умножении матрицы G'_{σ_1,σ_2} на Q мы считаем скалярное произведение всех строк первой матрицы на столбцы второй.

1. Любая строка g_i известных подматриц $G(r-1, m-2)$ при умножении на столбец q^j матрицы Q даст 0 или 1, так как скалярное произведение векторов $g_i \cdot q^{jT}$ будет являться некоторой линейной комбинацией элементов столбца q^j. По определению перестановочной матрицы в любом столбце и строке существует только один ненулевой элемент.

Значит во всех элементах строк с номерами $k \in [\sum_{i=0}^{r}\binom{m-2}{i}+1, \sum_{i=0}^{r}\binom{m}{i} - \sum_{i=0}^{r-2}\binom{m-2}{i}]$ в новой матрице будут находиться константы, $d = 0$.

2. При перемножении первых $k \in [1, \sum_{i=0}^{r}\binom{m-2}{i}]$ и последних $\sum_{i=0}^{r-2}\binom{m-2}{i}$ строк матрицы G'_{σ_1,σ_2} при перемножении строк соответсвующим $\Omega_{f_i} \sim a_i$ и $\Omega_{f_i} \sim b_i$ будут получаться мономы степени $d = 2$. Из строения матрицы G'_{σ_1,σ_2} видно, что таких мономов будет

$n \cdot n \cdot (m-2)$ a_i и $n/4 \cdot n \cdot (m-2)$ b_i. Продолжая аналогично рассуждения для строк, соответствующих $a_i a_j, ..., a_1 ... a_{r-2}$ и $b_i b_j, ..., b_1 ... b_{r-2}$, получим формулу для количества всех мономов степеней $d \leq r-1$

$$n^2 \binom{m-2}{i} + \frac{n^2}{4} \binom{m-2}{i} = \frac{5}{4} \sum_{i=1}^{r-2} n^2 \binom{m-2}{i-1}.$$

3. Так как последний блок матрицы G'_{σ_1,σ_2} – $G(r-2, m-2)$ мономов степени большей $r-1$ при умножении на перестановочную матрицу давать не может, то для степеней $d = r, r+1$ учитываем только строки из первого блока, соответствующие $a_1 ... a_{r-1}, ..., a_1 ... a_r$. Они дадут $n^2(\binom{m-2}{r-1} + \binom{m-2}{r})$ мономов степени $d = r, r+1$, а значит и новых переменных.

4. Аналогичное рассмотрение случая $r = 1$ дает соответствующую формулу. □

2) II этап: Сведение линейной системы к КНФ: После первого этапа каждый полином был приведен к «длинным» суммам по модулю 2, то есть может быть представлен как $x_1 \oplus x_2 \oplus ... \oplus x_l = 0$. Однако для «длинных» сумм невозможно построить КНФ, [20], так как происходит экспоненциальный рост количества ЭД с ростом количества переменных. Чтобы этого избежать, необходимо разбить все «длинные XOR» на более короткие, например, длины 4.

Так, уравнение

$$x_1 \oplus x_2 \oplus ... \oplus x_l = 0 \tag{17}$$

можно привести к эквивалентной системе уравнений

$$\begin{pmatrix} x_1 \oplus x_2 \oplus x_3 \oplus y_1 & = & 0 \\ y_1 \oplus x_4 \oplus x_5 \oplus y_2 & = & 0 \\ \vdots & & \vdots \\ y_i \oplus x_{4i+2} \oplus x_{4i+3} \oplus y_{i+1} & = & 0 \\ \vdots & & \vdots \\ y_h \oplus x_{l-2} \oplus x_{l-1} \oplus x_l & = & 0 \end{pmatrix}, \tag{18}$$

если l - четное. Если же нечетное, то последнее уравнение будет длины 3. Здесь $h = \lceil l/2 \rceil - 2$ - количество новых переменных. Система будет состоять из $h+1$ уравнения, каждое из которых породит 8 ЭД в КНФ длины 4.

Утверждение 3. *Существует полиномиальный алгоритм, который позволяет получать из нелинейной системы булевых уравнений*

$$(G'_{\sigma_1,\sigma_2} \cdot Q^T) \cdot H_{pub}^T = 0.$$

эквивалентную ФАЛ, представленную своей КНФ, длина которой будет не больше чем

$$L_{12} = 4n^3 k + n^2 \left[\sum_{i=1}^{r-2} \frac{5(i+2)}{4} \binom{m-2}{i} + \right.$$
$$\left. + (r+1) \binom{m-2}{r-1} + (r+2) \binom{m-2}{r} - 4k^2 \right] - 8nk + 8k^2,$$

а количество переменных не будет превышать

$$N_{12} = n^3 \frac{k}{2} + n^2 \left[1 + \frac{5}{4} \sum_{i=1}^{r-2} \binom{m-2}{i-1} + \binom{m-2}{r-1} \right.$$
$$\left. + \binom{m-2}{r} - \frac{k^2}{2} \right] + n \left(\frac{m}{2} - 2k - 1 \right) + 2k^2$$

то есть параметры получаемой КНФ кубически зависят от длины входа n.

Доказательство. Рассмотрим получившуюся в разделе IV.C нелинейную систему

$$(G'_{\sigma_1,\sigma_2} \cdot Q^T) \cdot H_{pub}^T = 0.$$

из $k \times (n-k)$ уравнений.

Степень любого из мономов не превосходит $r+1$ (следует из построения), длина уравнений не превосходит $2^m \cdot 2^m = 2^{2m} = n^2$ (длина уравнения, являющегося элементом матрицы $G'_{\sigma_1,\sigma_2} \cdot Q^T$ не превышает $2^{m-2} \cdot 4 = 2^m$, в матрице же H_{pub}: $2^m = n$ строк), а количество неизвестных равно $N = \frac{n(m-2)}{2} + n^2$.

Согласно I-II этапам сведения полиномиальной системы к линейной и далее линейной системы к КНФ, описанным выше, проведем преобразования этой системы, чтобы получить КНФ:

1. Все мономы степени $d > 1$ заменяются на новый переменные, их кол-во не превышает N_0 (16) из Утверждения 2. Для каждой новой переменной количество новых ЭД будет равно произведению $(d+1)$ на количество таких переменных, соответствующих моному степени d. Просуммировав по всем возможным $d = 3, ..., r+1$ получаем формулу $L = \frac{5n^2}{4} \sum_{i=1}^{r-2} (i+2) \binom{m-2}{i} + n^2((r+1)\binom{m-2}{r-1} + (r+2)\binom{m-2}{r})$ новых ЭД. После этого получается система из $k \times (n-k)$ уравнений.

Количество переменных на данном этапе $N = \frac{n(m-2)}{2} + n^2$ (сколько было изначально) $+N_0$.

2. Все «длинные» суммы по модулю 2 (17) должны быть заменены на системы уравнений (18), состоящих из меньшего количества слагаемых (допустим, 4).

Каждое уравнение длины l (17) породит $h = \frac{l}{2} - 2$ новых переменных и $h+1$ уравнение (18). А каждое укороченное уравнение заменится на КНФ, состоящее из 8 ЭД длины 4. Получаем:

Длина каждого уравнения $l \leq n^2$, следовательно, добавится $\leq \frac{n^2}{2} - 2$ новых переменных для каждого уравнения и $8(h+1) \leq 4n^2 - 8$ ЭД.

Значит всего добавится не больше $N_{xor} = k \cdot (n-k) \cdot (\frac{n^2}{2} - 2)$ переменных.

Длина КНФ при этом увеличится на количество уравнений умноженное на количество новых ЭД, т.е. $L_{xor} = k \times (n-k) \cdot 8(h+1) \leq k(n-k)(4n^2 - 8)$.

3. Таким образом, получаем, что общее число переменных не будет превосходить (при $r > 1$):

$$N_{12} = N + N_0 + Nxor$$

$$N_{12} = \frac{n(m-2)}{2} + n^2 + n^2 \left[\frac{5}{4} \sum_{i=1}^{r-2} \binom{m-2}{i-1} + \right.$$
$$\left. + \binom{m-2}{r-1} + + \binom{m-2}{r} \right] + k \cdot (n-k) \cdot (\frac{n^2}{2} - 2),$$

$$N_{12} = n^3\frac{k}{2} + n^2\left[1 + \frac{5}{4}\sum_{i=1}^{r-2}\binom{m-2}{i-1} + \binom{m-2}{r-1} + \right.$$
$$\left. + \binom{m-2}{r} - \frac{k^2}{2}\right] + n(\frac{m}{2} - 2k - 1) + 2k^2$$

а количество ЭД в КНФ, получаеющейся на этом этапе, не превосходит

$$L_{12} = L + L_{xor}$$

$$L_{12} = \frac{5n^2}{4}\sum_{i=1}^{r-2}(i+2)\binom{m-2}{i} +$$
$$+ n^2\left[(r+1)\binom{m-2}{r-1} + (r+2)\binom{m-2}{r}\right] +$$
$$+ k(n-k)(4n^2-8)$$

$$L_{12} = 4n^3k + n^2\left[\sum_{i=1}^{r-2}\frac{5(i+2)}{4}\binom{m-2}{i}\binom{m-2}{i} + \right.$$
$$\left. + (r+1)\binom{m-2}{r-1} + (r+2)\binom{m-2}{r} - 4k^2\right]$$
$$- 8nk + 8k^2$$

□

F. Сведение общей задачи выполнимости КНФ к задаче выполнимости КНФ для схемы ЭЦП $pqsigRM$

В процессе сведения полиномиальной системы к КНФ была получена КНФ, однако, пока что для общей задачи. Мы ввели неизвестные подматрицы следующим образом:

$$G(r, m-2)^{\sigma_1} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ a_1^1 & a_2^1 & \dots & a_{2^{m-2}}^1 \\ \vdots & \vdots & & \vdots \\ a_1^{m-2} & a_2^{m-2} & \dots & a_{2^{m-2}}^{m-2} \\ a_1^1a_1^2 & a_2^1a_2^2 & \dots & a_{2^{m-2}}^1a_{2^{m-2}}^2 \\ \vdots & \vdots & & \vdots \\ a_1^1...a_1^r & a_2^1...a_2^r & \dots & a_{2^{m-2}}^1...a_{2^{m-2}}^r \end{pmatrix}$$

Нигде в этом условии не было учтено, что все векторы $a_i = (a_i^1, ..., a_i^{m-2}), i \in [1, 2^{m-2}]$ должны быть попарно различны.

Утверждение 4. *При введении дополнительных условий на неизвестные векторы $a_i = (a_i^1, ..., a_i^{m-2}), b_i = (b_i^1, ..., b_i^{m-2})$, $i \in [1, 2^{m-2}]$ матриц $G(r, m-2)^{\sigma_1}$ и $G(r-2, m-2)^{\sigma_2}$ соответственно общее количество новых ЭД, добавленных в построенную на предыдущих этапах КНФ, равно $L_{gen} = \frac{n^2}{32}\left(\frac{n}{4} - 1\right)$, а новых переменных при этом не прибавится.*

Доказательство. Пусть есть векторы $x = (x_1, ..., x_w)$ и $y = (y_1, ..., y_w)$. Тогда условию $x \neq y$ будет эквивалентна следующая формула $\bigvee_{i=1}^{m} x_i \oplus y_i = 1$. Рассмотрим булеву функцию $f(x, y) = (x_1 \oplus y_1) \vee ... \vee (x_w \oplus y_w)$ от $2m$ переменных, с возможным количеством значений 2^{2m}. Вектора x, y будут совпадать на 2^m наборах, то есть у функции f будет 2^m нулей.

В КНФ этой функции все ЭД будут длины $2m$, так как иначе существовали бы соседние наборы $x'_1, ..., x'_w$ и $y'_1, ..., y'_w$, на которых $f(x', y') = 0$, но это невозможно. Значит для f КНФ является СКНФ и количество ЭД в ней равно $n = 2^m$.

То есть каждая такая функция, соотвествующая паре векторов a_i, a_j (частям столбца подматрицы $G(r, m-2)^{\sigma_1}$, породит n ЭД. Количество таких пар будет равно $\frac{n'(n'-1)}{2}, n' = 2^{m-2} = \frac{n}{4}$. Значит общее количество новых ЭД будет равно $L_{gen} = \frac{n'^2(n'-1}{2} = \frac{n^2(n/4-1)}{32} = \frac{n^3}{128} - \frac{n^2}{32}$. Новых переменных не добавляется. □

Также никак не было учтено, что матрица Q (**??**) является перестановочной, т.е. что в каждом столбце и строке есть только один ненулевой элемент.

Утверждение 5. *При добавлении условий перестановочную матрицу Q количество переменных в КНФ не меняется, а количество ЭД увеличится на $L_Q = n^2(n-1) + 2n$.*

Доказательство. Очевидно, что количество переменных не меняется, уравнений увеличивается на $2n$. Каждое уравнение $q_1 + ... + q_n = 1$ заменяется булевой формулой вида $(\bar{q}_1 \vee \bar{q}_2)(\bar{q}_2 \vee \bar{q}_3)...(\bar{q}_{n-1} \vee \bar{q}_n)(q_1 \vee q_2 \vee ... \vee q_n) = 1$, где сначала берутся все возможные пары из $q_1, ..., q_n$. Поэтому каждое уравнение породит $\binom{n}{2} + 1$ ЭД. То есть всего $L_Q = 2n \cdot (\binom{n}{2} + 1) = 2n \cdot (\frac{n!}{(n-2)!2!} + 1) = 2n \cdot (\frac{n(n-1)}{2} + 1) = n^2(n-1) + 2n.$ новых ЭД

□

Таким образом, из утверждений 4-5 очевидно следует утверждение:

Утверждение 6. *Существует полиномиальный алгоритм сведения задачи нахождения секретного ключа схемы ЭЦП $pqsigRM$, основанной на модифицированных кодах Рида-Маллера, к задаче ВЫПОЛНИМОСТЬ КНФ, длина которой не превышает*

$$L = L_{12} + L_{gen} + L_Q \iff$$

$$L = n^3(4k + 1\frac{1}{2^7}) +$$
$$n^2\left[\sum_{i=1}^{r-2}\frac{5(i+2)}{4}\binom{m-2}{i} + (r+1)\binom{m-2}{r-1} + \right.$$
$$\left. + (r+2)\binom{m-2}{r} - 4k^2 - 1\frac{1}{2^5}\right]$$
$$+ 2n(1-4k) + 8k^2,$$

то есть длина не превосходит полинома 3 степени от $n = 2^m$, а количество переменных не превосходит

$$N = N_{12} = n^3\frac{k}{2} + n^2\left[\frac{5}{4}\sum_{i=1}^{r-2}\binom{m-2}{i-1} + \right.$$
$$\left. + \binom{m-2}{r-1} + \binom{m-2}{r} + 1 - \frac{k^2}{2}\right]$$
$$+ n(\frac{m}{2} - 2k - 1) + 2k^2,$$

то есть тоже не превосходит полинома 3 степени от n.

Теорема IV.1. *Существует полиномиальный алгоритм сведения задачи криптоаналитика нахождения секретного ключа криптосистемы $pqsigRM$, основанной на модифицированных кодах Рида-Маллера, к задаче ВЫПОЛНИМОСТЬ КНФ. При этом длина получающейся КНФ*

$$L \leq 2n^3(2k+1) + \frac{5}{4}n^2 \log_2 n + 2(n+4k^2)$$

а количество переменных в этой КНФ

$$N \leq n^3\frac{k}{2} + n^2\frac{5k}{4} + \frac{n \cdot \log_2 n}{2} + 2k^2$$

G. Программная реализация

Полученный алгоритм сведения задачи криптоаналитика поиска секретных ключей к задаче ВЫПОЛНИМОСТЬ КНФ был реализован программно на Python.

Прямая реализация описанного выше алгоритма была слишком медленной. Уже на параметрах $r = 4, m = 7$ она работала больше 10 часов. Поэтому была произведена оптимизация (одновременное выполнение некоторых шагов алгоритма, отказ от лишних проверок и т.д.).

Результаты работы программы для различных параметров (r, m) приведена в Таблицах I, II:

	(1,4)	(2,4)	(3,4)	(1,5)	(2,5)	(3,5)
N	776	1032	1168	4120	7192	9008
L	5432	5944	6352	41248	47392	51968
Time(sec)	0,50	0,47	0,34	5,33	7,2	4,53
Space(Mb)	0,106	0,126	0,140	0,776	1,0	1,2

Таблица I

Результаты работы программной реализации алгоритма, m=4, m=5

	(2,6)	(3,6)	(4,6)	(5,7)	(6,8)
N	45120	65664	75904	627008	5112576
L	358400	405376	425856	3435776	27554304
Time(sec)	101,3	83,8	42,9	6min55sec	3h 40min
Space(Mb)	8,5	10,7	11,7	105,1	895,8

Таблица II

Результаты работы программной реализации алгоритма

Из полученых данных на количество переменных и ЭД видно, что алгоритм при фиксированном m будет работать эффективнее значениях r близких к m, так как время создания КНФ в формате DIMACS при большем r значительно меньше, несмотря на то, что объем получаемого файла растет еще и засчет быстрого роста размеров комментариев – описания переменных.

V. Результаты экспериментов

В разделе IV был сформулирован и исследован полиномиальный алгоритм сведения задачи криптоаналитика поиска секретного ключа по открытому к задаче Выполнимости КНФ. Этот алгоритм был реализован программно на языке Python.

Далее на данных, полученных с помощью этой программной реализации, были запущены некоторые победители конкурсов SAT Competition 2018, SAT Race 2019, а также решатели, считающиеся эффективными, но не участвующие в конкурсе. Результаты этих экспериментов приведены в Таблицах III-VI. Время измерялось с помощью утилиты time, итоговый результат рассчитывался как $t = user + sys$ в секундах.

	(1,4)	(2,4)	(3,4)
CaDiCal	0,009	0,014	0,019
MapleLCMDiscChronoBT-DL-v3	0,017	0,041	0,041
Maple_LCM_Scavel	0,017	0,017	0,020
Maple_CM_ordUIP	0,026	0,038	0,055
Painless-v2	0,064	0,08	0,079
Plingeling	0,063	0,074	0,09
abcdSAT	0,117	0,11	0,129
picosat	0,008	0,012	0,012
picomus	0,017	0,021	0,021
zchaff64	0,01	0,014	-
slime	0,028	0,025	0,027

Таблица III

Результаты работы SAT-решателей для задачи криптоаналитики получения секретных ключей схемы $pqsigRM$, m=4

	(1,5)	(2,5)	(3,5)
CaDiCal	0,037	0,040	0,043
MapleLCMDiscChronoBT-DL-v3	0,088	0,102	0,108
Maple_LCM_Scavel	0,109	0,097	0,099
Maple_CM_ordUIP	0,106	0,136	0,138
Painless-v2	0,247	0,279	0,318
Plingeling	0,153	0,169	0,172
abcdSAT	0,678	0,697	0,725
picosat	0,038	0,056	0,075
picomus	0,092	0,107	0,12
zchaff64	0,042	0,072	0,058
slime	0,168	0,19	0,194

Таблица IV

Результаты работы SAT-решателей для задачи криптоаналитики получения секретных ключей схемы $pqsigRM$, m=5

	(2,6)	(3,6)	(4,6)
CaDiCal	0,258	0,475	0,530
MapleLCMDiscChronoBT-DL-v3	1,417	1,631	1,746
Maple_LCM_Scavel	1,408	1,662	1,759
Maple_CM_ordUIP	1,901	2,140	2,184
Painless-v2	10,23	11,653	12,180
Plingeling	1,3	1,715	1,122
abcdSAT	24,323	24,807	25,643
picosat	0,26	0,423	0,441
picomus	0,877	1,248	1,341
zchaff64	0,51	0,78	0,829
slime	3,995	4,313	4,343

Таблица V

Результаты работы SAT-решателей для задачи криптоаналитики получения секретных ключей схемы $pqsigRM$, m=6

	(5,7)	(6,8)
CaDiCal	7,906	2min 13 sec
MapleLCMDiscChronoBT-DL-v3	35,063	11min 11sec
Maple_LCM_Scavel	34,65	11min 13sec
Maple_CM_ordUIP	37,047	-
Painless-v2	3min39sec	-
Plingeling	5,879	59,987
abcdSAT	36,626	36,484
picosat	12,343	3min 26sec
picomus	20,673	-
zchaff64	11,455	11,118
slime	1min 34sec	42min 14sec

Таблица VI

Результаты работы SAT-решателей для задачи криптоаналитики получения секретных ключей схемы $pqsigRM$, m=7, m=8

В среднем лучшие результаты показали решатели zchaff64, Plingeling и picomus. Решатель zchaff64, несмотря на то, что являлся победителем конкурса решателей

только 2007 года и после этого призовых мест вообще не занимал, показал в среднем наименьшее время решения задачи Выполнимости КНФ для задачи криптоаналитика.

Также решатели Plingeling, abcdSAT, Painless-v2 несмотря на то, что являются параллельными, на некоторых параметрах отработали хуже, чем обычные решатели. Связано это может быть с тем, что такие программы задействуют планировщики для распределения задач по ядрам, и они должны быть определенным образом настроены. Поэтому на домашнем компьютере параллельные решатели могут работать даже хуже.

VI. Заключение

Для новой схемы ЭЦП, основанной на модифицированных кодах Рида-Маллера, найден эффективный метод атаки для некоторых значений параметров. Предложен полиномиальный алгоритм сведения задачи криптоаналитика поиска секретных ключей криптосистемы $pqsigRM$ к задаче ВЫПОЛНИМОСТИ КНФ с параметрами:

1) Количество переменных

$$N \leq n^3\frac{k}{2} + n^2\frac{5k}{4} + \frac{n \cdot \log_2 n}{2} + 2k^2;$$

2) Количество ЭД (длина КНФ)

$$L \leq 2n^3(2k+1) + \frac{5}{4}n^2 \log_2 n + 2(n + 4k^2).$$

Алгоритм сведения реализован программно на Python и оптимизирован для более быстрой работы. Алгоритм протестирован на небольших значениях параметров (r, m). Проведены эксперименты по запуску лучших SAT-решателей последних лет на данных, полученных с помощью реализации алгоритма сведения. Схема для всех рассматриваемых параметров была взломана SAT-решателями, то есть получено решение задачи ВЫПОЛНИМОСТИ КНФ, а значит получены секретные ключи.

Библиография

[1] P. W. Shor, «Algorithms for Quantum Computation: Discrete Logarithms and Factoring», SFCS '94, с. 124—134, 1994. doi: 10.1109/SFCS.1994.365700.

[2] F. Arute, K. Arya, Babbush и et al, «Quantum supremacy using a programmable superconducting processor», *Nature*, № 574, с. 505—510, 2019, https://doi.org/10.1038/s41586-019-1666-5.

[3] National Academies of Sciences Engineering, and Medicine, *Quantum Computing: Progress and Prospects*, E. Grumbling и M. Horowitz, ред. Washington, DC: The National Academies Press, 2019. doi: 10 . 17226 / 25196. url: https : //www.nap.edu/catalog/25196/quantum-computing-progress-and-prospects.

[4] Л. Федичкин, «Квантовые компьютеры», *Наука и жизнь*, № 1, 2001.

[5] Т. Фолджер, «Квантовый взлом», *В мире науки*, № 4, с. 94—102, 2016.

[6] R. J. McEliece, «A public key cryptosystem based on algebraic coding theory», 1978.

[7] A. Barg, *Complexity Issues in Coding Theory*, 1997.

[8] Yuan Xing Li, R. H. Deng и Xin Mei Wang, «On the equivalence of McEliece's and Niederreiter's public-key cryptosystems», *IEEE Transactions on Information Theory*, т. 40, № 1, с. 271—273, 1994.

[9] В. Сидельников и С. Шестаков, «О системе шифрования, построенной на основе обобщенных кодов Рида-Соломона», *Дискретная математика*, т. 4, № 3, с. 57—63, 1992.

[10] В. Сидельников, «Открытое шифрование на основе двоичных кодов Рида-Маллера», *Дискретная математика*, т. 6, № 3, с. 3—20, 1994.

[11] L. Minder и A. Shokrollahi, «Cryptanalysis of the Sidelnikov Cryptosystem», EUROCRYPT '07, с. 347—360, 2007. doi: 10.1007/978-3-540-72540-4_20.

[12] М. Бородин и И. Чижов, «Эффективная атака на криптосистему Мак-Элиса, построенную на основе кодов Рида - Маллера», *Дискретная математика*, т. 26, с. 10—20, янв. 2014. doi: 10.4213/dm1264.

[13] Y. Lee, W. Lee, Y.-S. Kim и J.-S. No, *A Modified pqsigRM: RM Code-Based Signature Scheme*, https://eprint.iacr.org/2019/678, 2019.

[14] N. T. Courtois, M. Finiasz и N. Sendrier, «How to Achieve a McEliece-Based Digital Signature Scheme», в *Advances in Cryptology — ASIACRYPT 2001*, C. Boyd, ред., Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, с. 157—174.

[15] С. Кук, «Сложность процедур вывода теорем», *Кибернетический сборник*, № 12, с. 5—15, 1971.

[16] И. В. Чижов и М. А. Бородин, «Задача ВЫПОЛНИМОСТЬ и восстановление секретного ключа криптосистемы Мак-Элиса на основе кодов Рида-Маллера», в *Тихоновские чтения: Научная конференция, МГУ имени М.В. Ломоносова, 29-31 октября 2012 г.: Тезисы докладов, секция Вычислительная математика и кибернетика*, МАКС Пресс Москва, 2012, с. 131—133.

[17] Ф. Мак-Вильямс и Н. Слоэн, *Теория кодов, исправляющих ошибки*. 1979.

[18] I. Dumer, «Recursive decoding and its performance for low-rate Reed-Muller codes», *IEEE Transactions on Information Theory*, т. 50, № 5, с. 811—823, 2004. doi: 10.1109/TIT.2004.826632.

[19] Gregory V. Bard, Nicolas, T. Courtois и Chris Jefferson, *Efficient Methods for Conversion and Solution of Sparse Systems of Low-Degree Multivariate Polynomials over GF(2) via SAT-Solvers*, https://eprint.iacr.org/2007/024, 2007.

[20] N. Creignou и H. Daude, «Satisfiability threshold for random XOR-CNF formulas», *Discrete Applied Mathematics*, т. 96-97, с. 41—53, окт. 1999. doi: 10.1016/S0166-218X(99)00032-3.

Logical cryptanalysis as part of strength research into particular code-based signature scheme

Ivan V. Chizhov, Nataliia P. Tashevtseva

***Abstract*—The authors propose an algorithm, which converts input for the cryptoanalyst problem of revealing secret keys of Code-Based Signature Scheme pqsigRM to an equal input for the SATISFIABILITY problem. It is proved in the paper, that the proposed attack is polynomial. The set of parameters of resulting CNF – length and the number of used variables, are theoretically assessed. The practical implementation of the proposed algorithm on Python is developed, which effectively creates the desired CNF in DIMACS format based on arbitrary pgsigRM scheme parameters r, m. Furthermore, the article contains experiment results, including execution results of the designed program for some values of r and m and performance of several open-source SAT-solvers, winners of SAT Competition 2018 and SAT Race 2019 combined with other solvers used earlier for McEliece cryptosystem analysis, on solving the satisfiability problem for the resulting CNF for some values of r, m parameters of original cryptoanalyst problem. A set of parameters for which attack can be applied, that is the secret keys can be recovered, is described.**

***Keywords*—Post-quantum cryptography, polynomial attack, digital signature, Reed-Muller codes**

References

[1] P. W. Shor, «Algorithms for quantum computation: Discrete logarithms and factoring», SFCS '94, pp. 124–134, 1994. doi: 10.1109/SFCS.1994.365700.

[2] F. Arute, K. Arya, Babbush, and et al, «Quantum supremacy using a programmable superconducting processor», *Nature*, no. 574, pp. 505–510, 2019, https://doi.org/10.1038/s41586-019-1666-5.

[3] National Academies of Sciences, Engineering, and Medicine, *Quantum Computing: Progress and Prospects*, E. Grumbling and M. Horowitz, Eds. Washington, DC: The National Academies Press, 2019. doi: 10 . 17226 / 25196. [Online]. Available: https : / / www . nap . edu / catalog / 25196 / quantum - computing-progress-and-prospects.

[4] L. Fedichkin, «Kvantovye komp'yutery [Quantum computers]», *Nauka i zhizn' [Science and Life]*, no. 1, 2001.

[5] T. Folger, «Kvantovyy Vzlom [Quantum Hack]», *V mire nauki [In the world of science]*, no. 4, pp. 94–102, 2016.

[6] R. J. McEliece, «A public key cryptosystem based on algebraic coding theory», 1978.

[7] A. Barg, *Complexity Issues in Coding Theory*, 1997.

[8] Yuan Xing Li, R. H. Deng, and Xin Mei Wang, «On the equivalence of McEliece's and Niederreiter's public-key cryptosystems», *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 271–273, 1994.

[9] V. Sidelnikov and S. Shestakov, «O sisteme shifrovaniya, postroennoy na osnove obobshchennykh kodov Rida-Solomona [On an encoding system constructed on the basis of generalized Reed-Solomon codes]», *Diskretnaya matematika [Discrete mathematics]*, vol. 4, no. 3, pp. 57–63, 1992.

[10] V. Sidelnikov, «Otkrytoe shifrovanie na osnove dvoichnykh kodov Rida–Mallera [A public key cryptosystem based on Reed-Muller binary codes]», *Diskretnaya matematika [Discrete mathematics]*, vol. 6, no. 3, pp. 3–20, 1994.

[11] L. Minder and A. Shokrollahi, «Cryptanalysis of the Sidelnikov Cryptosystem», EUROCRYPT '07, pp. 347–360, 2007. doi: 10.1007/978-3-540-72540-4_20.

[12] M. A. Borodin and I. V. Chizhov, «Effektivnaya ataka na kriptosistemu Mak-Elisa, postroennuyu na osnove kodov Rida - Mallera [Effective attack on the McEliece cryptosystem based on Reed–Muller codes]», *Diskretnaya matematika [Discrete mathematics]*, vol. 26, pp. 10–20, Jan. 2014. doi: 10 . 4213 / dm1264.

[13] Y. Lee, W. Lee, Y.-S. Kim, and J.-S. No, *A Modified pqsigRM: RM Code-Based Signature Scheme*, Cryptology ePrint Archive, Report 2019/678, https://eprint.iacr.org/2019/678, 2019.

[14] N. T. Courtois, M. Finiasz, and N. Sendrier, «How to Achieve a McEliece-Based Digital Signature Scheme», in *Advances in Cryptology — ASIACRYPT 2001*, C. Boyd, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 157–174.

[15] S. A. Cook, «Slozhnost' protsedur vyvoda teorem [The complexity of theorem-proving procedures]», *Kiberneticheskiy sbornik [Cybernetic collection]*, no. 12, pp. 5–15, 1971.

[16] I. V. Chizhov and M. A. Borodin, «Zadacha VYPOLNIMOST" i vosstanovlenie sekretnogo klyucha kriptosistemy Mak-Elisa na osnove kodov Rida-Mallera [SATISFIABILITY problem and recovering of secret keys of McElice cryptosystem based on Reed-Muller codes]», Russian, in *Tikhonovskie chteniya: Nauchnaya konferentsiya, MGU imeni M.V. Lomonosova, 29-31 oktyabrya 2012 g.: Tezisy dokladov, sektsiya Vychislitel'naya matematika i kibernetika [Tikhonov readings: Science conference, MSU named after M.V.Lomonosov, 29-31 of October 2012: Report abstracts, Computational mathematics and cybernetics]*, MAKS Press Moskva [MAKS Press Moscow], 2012, pp. 131–133.

[17] F. MacWilliams and N. Sloane, *Teoriya kodov, ispravlyayushchikh oshibki [The Theory of Error-Correcting Codes]*. 1979.

[18] I. Dumer, «Recursive decoding and its performance for low-rate Reed-Muller codes», *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 811–823, 2004. doi: 10.1109/TIT.2004.826632.

[19] G. V. Bard, N. T. Courtois, and C. Jefferson., *Efficient Methods for Conversion and Solution of Sparse Systems of Low-Degree Multivariate Polynomials over GF(2) via SAT-Solvers*, https://eprint.iacr.org/2007/024, 2007.

[20] N. Creignou and H. Daude, «Satisfiability threshold for random XOR-CNF formulas», *Discrete Applied Mathematics*, vol. 96-97, pp. 41–53, Oct. 1999. doi: 10.1016/S0166-218X(99)00032-3.