ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 58 **2022** **Вып. 2**

УДК 621.391 : 004.056.5 : 519.725

© 2022 г. **В.В. Зяблов, Ф.И. Иванов, Е.А. Крук, В.Р. Сидоренко**[1]

О НОВЫХ ЗАДАЧАХ В АСИММЕТРИЧНОЙ КРИПТОГРАФИИ, ОСНОВАННОЙ НА ПОМЕХОУСТОЙЧИВОМ КОДИРОВАНИИ[2]

Рассматривается задача построения криптосистем с открытым ключом на основе помехоустойчивых кодов. Данный класс криптосистем на сегодняшний день является устойчивым к атакам с использованием квантового компьютера и потому может быть отнесен к методам постквантовой криптографии. Основным недостатком кодовой криптографии является очень большая длина открытого ключа. Большинство усилий по преодолению этого недостатка сводилось к замене кода Гоппы, который использовался в исходной криптосистеме, на код из другого множества, позволяющего описать открытый ключ более компактно, при этом сохранив стойкость криптосистемы к различным атакам. Здесь предложен другой подход к сокращению длины ключа – мы ставим задачу простого описания множества исправимых кодом ошибок, вес которых превосходит половину его минимального расстояния или которые не могут быть исправлены без знания некоторого скрытого преобразования. Если структура кода позволяет дать такое описание множества ошибок, то сложность большинства атак на зашифрованный текст (например, атака по информационным совокупностям) существенно возрастает.

Ключевые слова: криптографическая система Мак-Элиса, декодирование по информационным совокупностям, обобщенные коды Рида – Соломона, постквантовая криптография.

DOI: 10.31857/S0555292322020077, **EDN:** DZRXPW

§ 1. Введение

Методы теории помехоустойчивого кодирования давно используются в криптографии. С их помощью были построены одни из первых криптосистем с открытым ключом [1] и системы цифровой подписи [2]. Однако в отличие от алгебраических криптосистем, основанных на задачах факторизации [3] и вычисления дискретного логарифма [4], кодовые криптосистемы фактически не применяются на практике. Хотя кодовые криптосистемы выигрывают у алгебраических по времени шифрования/дешифрования [5], их использование в значительной степени ограничивается рядом объективных и субъективных факторов.

Во-первых, алгебраические системы возникли несколько раньше кодовых и сразу прошли через многочисленные испытания их безопасности. Последнее обстоятельство, в условиях отсутствия доказательной стойкости криптосистем с открытым ключом, является определенной гарантией безопасности.

[1] Работа В.Р. Сидоренко выполнена при поддержке европейского исследовательского совета ERC в рамках инновационной программы “Горизонт 2020” (номер гранта 801434).

[2] В статье использованы результаты проекта “Разработка методов достоверной и целостной передачи информации в многопользовательских системах с использованием помехоустойчивых кодов и цифровых водяных знаков”, выполненного в рамках Программы фундаментальных исследований НИУ ВШЭ в 2021 г.

Во-вторых, для первых кодовых криптосистем (система Мак-Элиса) было характерно наличие открытого ключа, существенно более длинного, чем для алгебраических систем (например, системы RSA).

Дальнейшие исследования кодовых криптосистем позволили существенно уменьшить размер открытого ключа [6, 7], а разработка новых методов решения задачи факторизации [8] заставила увеличить длину открытого ключа алгебраических криптосистем настолько, что эти величины открытых ключей стали соизмеримы [9]. Однако алгебраические криптосистемы остаются и сегодня основным инструментом организации криптографической безопасности.

Такое положение начало меняться в последние годы в связи с появлением понятия постквантовой криптографии [10] и развитием ряда новых областей использования криптографических методов.

Активные исследования в области создания так называемого квантового компьютера, моделирующего вычислительные процессы на квантовом уровне, привело к построению алгоритмов, ориентированных на этот компьютер. Одним из главных достижений теории квантовых алгоритмов явилась разработка Шором полиномиального алгоритма решения задачи факторизации на квантовом компьютере [11]. Получение этого алгоритма означает, что после создания достаточно мощного квантового компьютера системы защиты, основанные на алгебраических криптосистемах (а они составляют абсолютное большинство), окажутся скомпрометированными. В связи с этим появилось понятие постквантовой криптографии, т.е. криптографии, стойкость которой не подвергнется сомнению в связи с появлением квантового компьютера. Задача декодирования линейных кодов, лежащая в основе кодовых криптосистем, является *NP*-трудной [12] и, по-видимому, не будет решена за полиномиальное время даже с помощью квантовых компьютеров.

Современная практика сенсорных сетей, облачных вычислений и ряда других направлений инфокоммуникационных технологий выдвигает задачу создания так называемой "легкой криптографии" – криптографических алгоритмов, обеспечивающих достаточный уровень безопасности при использовании устройств с ограниченными вычислительными ресурсами [13]. Для целей легкой криптографии криптосистемы, основанные на теории кодирования, оказываются более перспективными [14], чем алгебраические. Кодовые криптосистемы требуют меньшего числа операций и используют операции линейной алгебры, реализация которых предпочтительна по сравнению с арифметическими операциями.

Все это определило новый всплеск интереса к кодовой криптографии и, возможно, новый прикладной этап в ее развитии.

Как уже было отмечено выше, существует ряд попыток преодолеть главный недостаток кодовых криптосистем – большую длину открытого ключа. Основная идея этих улучшений состоит в замене двоичного кода Гоппы, который используется в исходной криптосистеме Мак-Элиса, на какой-то другой с определенной структурой, позволяющей уменьшить размер открытого ключа. Например, в работе [6] коды Гоппы заменены подкодами квазициклических обобщенных кодов Рида – Соломона. Это позволяет получить криптосистему с размером ключа от 6000 до 11000 бит и уровнем безопасности от 2^{80} до 2^{107}. В [15] предлагается использовать квазициклические коды с умеренной плотностью проверок (QC-MDPC). Это приводит к значительному уменьшению размера ключа до 0,6 КБайт, что делает криптосистему на основе таких кодов практически осуществимой. Очень похожая криптосистема, основанная на квазициклических кодах с малой плотностью проверок (QC-LDPC), была предложена в [16].

Главный недостаток замены кодов Гоппы кодами QC-LDPC или QC-MDPC заключается в том, что практически реализуемые алгоритмы их итеративного декодирования не гарантируют исправления ошибок заданного веса t даже при сравнитель-

но небольших значениях данной величины. Более того, практически используемые коды QC-LDPC или QC-MDPC обычно имеют сравнительно небольшое минимальное расстояние (порядка десятков для кодов скорости $R = 1/2$ и длины в несколько тысяч).

В данной статье мы предлагаем иной подход к выбору кода, который будет использован в качестве компоненты кодовой криптосистемы: вместо задачи компактного описания открытого ключа за счет структурности кода мы поставим задачу выбора тройки $(C_0, \mathcal{E}, \varphi)$, где C_0 – секретный (n, k, d)-код, $\mathcal{E}$ – множество ошибок, вносимых на этапе шифрования, а φ – преобразование полиномиальной сложности, отображающее множество вносимых ошибок в множество ошибок, исправимых кодом. Таким образом, ставится задача описания множества ошибок (не обязательно малого веса), исправимых кодом C_0. Структура данного кода должна быть спрятана от криптоаналитика. Сокращение длины открытого ключа в этом случае будет достигаться за счет того, что классические атаки на зашифрованный текст (например, атака по информационным совокупностям [17]) столкнутся со значительно более сложной, нежели задача исправления ошибок веса до $\dfrac{d-1}{2}$, задачей исправления ошибок веса, большего чем $\dfrac{d-1}{2}$. Это, в свою очередь, позволит перейти к значительно более коротким кодам с сохранением при этом требуемой сложности атаки.

§ 2. Криптосистема Мак-Элиса

Первой кодовой криптосистемой была система Мак-Элиса, предложенная в 1978 году в [1]. В качестве открытого ключа в ней использовалась двоичная матрица $\boldsymbol{G}$ размера $k \times n$, представляющая собой произведение матриц

$$\boldsymbol{G} = \boldsymbol{S}\boldsymbol{G}_0\boldsymbol{P}, \tag{1}$$

где $\boldsymbol{S}$ – невырожденная двоичная $(k \times k)$-матрица, $\boldsymbol{G}_0$ – порождающая матрица двоичного (n, k, d)-кода C_0, для которого известен "простой" (как правило, полиномиальный) алгоритм ξ декодирования ошибок кратности до половины кодового расстояния $t = \dfrac{d-1}{2}$, а $\boldsymbol{P}$ – перестановочная $(n \times n)$-матрица. Следует отметить, что матрица $\boldsymbol{S}\boldsymbol{G}_0$ порождает то же множество кодовых слов, что и $\boldsymbol{G}_0$, т.е. код C_0.

Шифрограмма для сообщения $\boldsymbol{x}$ в системе Мак-Элиса вычисляется следующим образом:

1. Генерируется случайный вектор $\boldsymbol{e}$ длины n из множества $\mathcal{E}_t$ векторов веса t;
2. Вычисляется шифрограмма

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{G} + \boldsymbol{e}. \tag{2}$$

Предполагается, что легальный получатель сообщения знает матрицы $\boldsymbol{S}$, $\boldsymbol{G}$, $\boldsymbol{P}$ в произведении (1), которые являются закрытым ключом криптосистемы. В этом случае легальный пользователь находит зашифрованное сообщение по следующему алгоритму:

1. Умножает $\boldsymbol{y}$ на $\boldsymbol{P}^{-1}$:

$$\boldsymbol{y}\boldsymbol{P}^{-1} = \boldsymbol{x}\boldsymbol{S}\boldsymbol{G}_0 + \boldsymbol{e}\boldsymbol{P}^{-1};$$

2. Декодирует полученный вектор кодом C_0 с порождающей матрицей $\boldsymbol{S}\boldsymbol{G}_0$. Поскольку вектор $\boldsymbol{e}\boldsymbol{P}^{-1}$ имеет вес t, то в результате декодирования будет получен вектор $\boldsymbol{x}\boldsymbol{S}$;
3. Получает $\boldsymbol{x}$ как $\boldsymbol{x} = \boldsymbol{x}\boldsymbol{S}\boldsymbol{S}^{-1}$.

Идея системы Мак-Элиса состоит в том, что после применения обратного преобразования $\boldsymbol{P}^{-1}$ к шифрограмме $\boldsymbol{y}$ вектор ошибки $\boldsymbol{e}\boldsymbol{P}^{-1}$ не меняет своего веса,

т.е. принадлежит тому же множеству векторов, которому принадлежал исходный вектор ошибки $\boldsymbol{e}$, используемый при шифровании. Поэтому для дешифрования легальному пользователю не надо было знать конкретный вектор ошибки – любой вектор ошибки веса до t декодировался в коде C_0 одним и тем же алгоритмом.

Открытым ключом системы Мак-Элиса является матрица $\boldsymbol{G}$. Действительно, стойкость описанной системы основана на сложности декодирования линейного кода C произвольной структуры. После преобразования матрицы, описываемого формулой (1), алгебраическая структура матрицы кода с простым декодированием ξ маскируется. Умножение справа на $\boldsymbol{P}$ переводит исходный код в эквивалентный код C, и простое декодирование ξ к нему применить нельзя.

§ 3. Атаки на кодовую криптосистему

В этом параграфе мы рассмотрим классификацию атак на кодовую криптосистему (безотносительно выбора конкретного кода, лежащего в основе криптосистемы).

Выделяют два основных типа атак на кодовую криптосистему: атака декодирования и структурная. Их основное отличие заключается в том, что атака декодирования направлена на извлечение зашифрованного сообщения $\boldsymbol{x}$ из шифротекста $\boldsymbol{y} = \boldsymbol{x}\boldsymbol{G} + \boldsymbol{e}$ путем декодирования $\boldsymbol{y}$ некоторым специально выбранным алгоритмом. Структурная атака направлена на восстановление секретного ключа $(\boldsymbol{S}, \boldsymbol{G}_0, \boldsymbol{P})$ из открытого ключа $\boldsymbol{G} = \boldsymbol{S}\boldsymbol{G}_0\boldsymbol{P}$ на основе некоторой доступной информации о структуре кода C_0 с порождающей матрицей $\boldsymbol{G}_0$. Отметим, что не всегда требуется найти первоначальное разложение $\boldsymbol{G} = \boldsymbol{S}\boldsymbol{G}_0\boldsymbol{P}$. Зачастую достаточным оказывается найти некоторое разложение $(\boldsymbol{S}', \boldsymbol{G}_0', \boldsymbol{P}')$, такое что $\boldsymbol{G} = \boldsymbol{S}'\boldsymbol{G}_0'\boldsymbol{P}'$, как это сделано, например, в атаке Сидельникова – Шестакова на криптосистему на основе обобщенного кода Рида – Соломона [18]. В данной статье мы не будем подробно останавливаться на анализе структурных атак применительно к предложенной нами криптосистеме, более детально фокусируясь на атаках по декодированию. Более того, нельзя гарантировать, что для заданной криптосистемы не найдется структурной атаки. Известно, что большинство успешных (полиномиальных по сложности атак) на различные варианты кодовых криптосистем относятся именно к классу структурных атак.

Атаки на основе декодирования подразделяются на следующие:

1. Атака на основе перебора информационных векторов – осуществляется путем перебора всех возможных векторов $\boldsymbol{x}$ до тех пор, пока не будет получено значение $\mathrm{wt}(\boldsymbol{x}\boldsymbol{G} - \boldsymbol{y}) = t$. Число попыток, необходимое для реализации данной атаки для (n, k)-кода C, оценивается сверху как $2^{\min(k, n-k)}$;
2. Атака на основе перебора векторов ошибки – осуществляется путем перебора всех возможных векторов $\boldsymbol{e}$ до тех пор, пока не будет получено $\mathrm{wt}\big((\boldsymbol{y} - \boldsymbol{e})\boldsymbol{H}^{\mathrm{T}}\big) = 0$, где $\boldsymbol{H}$ – проверочная матрица, соответствующая публичной порождающей матрице $\boldsymbol{G}$. Сложность данной атаки зависит от мощности множества ошибок, вносимых на этапе шифрования сообщения. Если в основе криптосистемы лежит двоичный (n, k)-код, исправляющий t ошибок, которые случайным образом вносятся на этапе шифрования, то среднее число попыток при данной атаке оценивается сверху величиной $\binom{n}{t}$;
3. Атака на основе поиска свободных от ошибок информационных совокупностей. Подробное описание данной атаки приведено в следующем параграфе.

Следует отметить, что цель любой атаки – это поиск простого декодирования в классе эквивалентных кодов, если известно, что по крайней мере для одного из кодов существует простое декодирование ξ. Разница двух подходов – структурного и на основе декодирования – состоит в том, что в случае атак на основе декодирования применяются общие методы декодирования линейных кодов с произвольной

структурой, а успешное применение структурной атаки позволяет использовать для декодирования полиномиальный декодер ξ.

3.1. Декодирование по информационным совокупностям. Задача декодирования по минимуму расстояния кода с произвольной структурой, как уже было отмечено во введении, является *NP*-трудной [12]. Декодирование ошибок кратности до t (до половины кодового расстояния), конечно, является относительно более простой, но тоже, по-видимому, экспоненциальной по сложности задачей, хотя доказательства ее *NP*-трудности на данный момент не представлено. Во всяком случае, полиномиальных алгоритмов решения этой задачи на данный момент не известно.

Далее мы дадим описание алгоритма декодирования по информационным совокупностям (ISD), на основе которого реализованы наиболее "перспективные" атаки на криптосистему Мак-Элиса.

Цель алгоритмов ISD – восстановить сообщение $\boldsymbol{x}$ из заданного вектора $\boldsymbol{y} = \boldsymbol{x}\boldsymbol{G}+\boldsymbol{e}$, где $\boldsymbol{G}$ – порождающая матрица (n,k)-кода C с минимальным расстоянием $d = 2t+1$ и $\mathrm{wt}(\boldsymbol{e}) \leqslant t$.

Пусть $\mathcal{I}$ является k-подмножеством набора координат $[n] := \{1,2,\ldots,n\}$, такое что $\mathcal{I}$ является информационной совокупностью кода C, а $\boldsymbol{G}_{\mathcal{I}}$ – подматрица $\boldsymbol{G}$, состоящая из столбцов с индексами из $\mathcal{I}$. Аналогично, пусть $\boldsymbol{e}_{\mathcal{I}}$ – вектор, состоящий из координат вектора $\boldsymbol{e}$ с индексами из $\mathcal{I}$.

Алгоритм декодирования ISD работает следующим образом:

1. Выбирается случайная информационная совокупность $\mathcal{I} \subset \{1,2,\ldots,n\}$;
2. Если $\mathrm{wt}(\boldsymbol{y}-\boldsymbol{y}_{\mathcal{I}}\boldsymbol{G}_{\mathcal{I}}^{-1}\boldsymbol{G}) \leqslant t$, то $\boldsymbol{y}_{\mathcal{I}}$ не содержит ошибок, а значит, $\mathrm{wt}(\boldsymbol{e}_{\mathcal{I}}) = 0$. Тогда $\boldsymbol{u} = \boldsymbol{y}_{\mathcal{I}}\boldsymbol{G}_{\mathcal{I}}^{-1}$. Иначе возврат к шагу 1.

Легко заметить, что вероятность P_k того, что заданная информационная совокупность не содержит ошибок, оценивается снизу как

$$P_k \leqslant \frac{\binom{n-t}{k}}{\binom{n}{k}} = \frac{\binom{n-k}{t}}{\binom{n}{t}}. \tag{3}$$

Это значит, что среднее число попыток поиска свободной от ошибок информационной совокупности не превышает $\dfrac{\binom{n}{t}}{\binom{n-k}{t}}$, что значительно меньше, нежели перебор по всем векторам ошибок. Поэтому для достижения требуемой стойкости криптосистемы Мак-Элиса приходится выбирать коды большой длины n. В частности, сам Мак-Элис предлагал использовать $(1024, 524, 101)$-код Гоппы, исправляющий $t = 50$ ошибок, для которого длина открытого ключа равна 536576 бит.

ISD-атака упоминалась еще в [1] и получила дальнейшее развитие в многочисленных публикациях (см., например, работу [19] и библиографию в ней). Существуют разные интерпретации и модификации исходного алгоритма ISD. Было предложено несколько различных улучшений, например, основанных на обобщенном парадоксе дней рождения. В работе [20] показано, что асимптотическая сложность декодирования по информационным совокупностям не превосходит $\widetilde{\mathcal{O}}(2^{0{,}0494n})$, что на данный момент является наилучшей известной оценкой.

3.2. Сложность атак на основе декодирования для криптосистемы Мак-Элиса. Приведем пример вычисления сложности атак. (*Здесь и далее под сложностью атаки мы будем понимать среднее число элементарных операций, необходимых для поиска сообщения $\boldsymbol{x}$ по зашифрованному сообщению $\boldsymbol{y}$*). Для классической крипто-

системы Мак-Элиса на основе $(1024, 524, 101)$-кода Гоппы, исправляющего $t = 50$ ошибок:

- Сложность атаки на основе перебора информационных векторов $2^{n-k}(n-k)k = 2^{518}$;
- Сложность атаки на основе перебора векторов ошибки $k(n-k)\binom{n}{t} = 524 \cdot 500 \cdot \binom{1024}{50} \approx 2^{302}$;
- Сложность атаки на основе поиска свободной от ошибки информационной совокупности оценивается сверху как

$$\frac{\binom{n}{t}}{\binom{n-k}{t}} k(n-k) = \frac{\binom{1024}{50}}{\binom{500}{50}} 524 \cdot 500 \approx 2^{72}.$$

Как уже было отмечено выше, сложность последней атаки оказалась наименьшей. Таким образом, стойкость (*под стойкостью криптосистемы мы будем понимать наименьшую сложность среди известных атак*) классической криптосистемы Мак-Элиса можно оценить минимумом сложности среди рассмотренных атак, который равен 2^{72}.

Далее будут обсуждаться способы увеличения сложности этой атаки путем модификации множества вносимых ошибок.

§ 4. Задача построения множества исправимых кодом ошибок

Все описанные в ранее указанных работах алгоритмы декодирования так или иначе опираются на то, что они исправляют "легкие" ошибки, вес которых значительно меньше длины кодового слова. Однако любой линейный код способен исправлять значительное количество ошибок большого веса. Пусть линейный код C_0 длины n задан над полем $\mathbb{F}$. Пространство $\mathbb{F}^n$ разобьем на смежные классы по C_0. Ясно, что код C_0 может исправлять только один вектор ошибки из каждого смежного класса, но это может быть любой вектор из этого класса. Это означает, что в формуле (2) можно использовать в качестве случайного маскирующего вектора (вектора ошибки $\boldsymbol{e}$) любые (не обязательно "легкие") векторы с условием, что они принадлежат разным смежным классам кода. При этом существенно возрастает сложность алгоритма декодирования по информационным совокупностям (она растет экспоненциально вместе с t), и для обеспечения требуемой стойкости системы можно использовать код существенно меньших размеров.

Однако здесь возникает другая задача, которая легко решалась (и даже не рассматривалась как задача) в исходной системе Мак-Элиса. Как генерировать множество вносимых ошибок, которые легальный пользователь будет декодировать с помощью кода C_0 с порождающей матрицей $\boldsymbol{G}_0$? Фактически мы должны задать множество ошибок $\mathcal{E}$, из которого будет случайным образом выбираться вектор ошибок, используемый в шифровании (2).

Вначале сформулируем свойства, которыми должно обладать это множество. Обозначим через $\mathcal{E}_0$ множество векторов ошибок $\boldsymbol{e}$, исправляемых кодом C_0 с помощью полиномиального алгоритма декодирования ξ, а через $\varphi(C_0)$ – некоторое обратимое преобразование кода (его порождающей матрицы, т.е. базиса). В результате этого преобразования получается код $C = \varphi(C_0)$. Пусть, кроме того, ω – требуемая стойкость криптосистемы.

Тогда мы можем сформулировать требования, предъявляемые к множеству $\mathcal{E}$:

(а) Существует алгоритм V генерации случайного вектора $\boldsymbol{e} \in \mathcal{E}$, имеющий полиномиальную сложность;

(b) Существует обратимое линейное преобразование $\varphi(C_0)$ полиномиальной сложности, отображающее код C_0 в код C;

(c) $\varphi^{-1}(\mathcal{E}) \subset \mathcal{E}_0$, причем $|\varphi^{-1}(\mathcal{E})| \geqslant \omega$;

(d) Для кода C_0 существует алгоритм ξ исправления ошибок из множества $\mathcal{E}$, имеющий полиномиальную сложность;

(e) Сложность декодирования ошибок из множества $\mathcal{E}$ в коде C не меньше чем ω, в частности, $|\mathcal{E}| \geqslant \omega$ – т.е. мощность множества должна препятствовать перебору по всем его элементам с целью вскрытия криптосистемы.

С учетом введенных обозначений предлагаемая обобщенная схема шифрования $\mathbb{S}$ может быть описана следующим образом:

- Публичный ключ в предлагаемой системе: порождающая матрица $\boldsymbol{G} = \varphi(\boldsymbol{G}_0)$ и алгоритм V;
- Секретный ключ: обратное преобразование φ^{-1}, матрица $\boldsymbol{G}_0$ и декодер ξ;
- Алгоритм шифрования:
  1. С помощью алгоритма V выбирается случайный вектор $\boldsymbol{e} \in \mathcal{E}$;
  2. По сообщению $\boldsymbol{x}$ вычисляется шифрограмма

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{G} + \boldsymbol{e}; \tag{4}$$

- Алгоритм дешифрования:
  1. Вычисляется $\boldsymbol{y}' = \varphi^{-1}(\boldsymbol{y}) = \boldsymbol{x}\varphi^{-1}(\boldsymbol{G}) + \varphi^{-1}(\boldsymbol{e}) = \boldsymbol{x}\boldsymbol{G}_0 + \boldsymbol{e}'$, где $\boldsymbol{e}' \in \mathcal{E}_0$;
  2. С помощью алгоритма ξ находится $\boldsymbol{x}$.

Хотя визуально задача декодирования (4) полностью совпадает с задачей (1), она должна быть сложнее за счет того, что исправление ошибок из $\mathcal{E}$ в коде C с произвольной структурой является более сложной задачей, чем исправление ошибок малой кратности. Этот факт следует непосредственно из того, что при заданных n, k вероятность P_k в (3) является монотонно убывающей функцией аргумента t, т.е. из $t < t' \leqslant \dfrac{n}{2}$ следует, что $P_k(t) \gg P_k(t')$.

Оценка эффективности кодовых криптосистем. Стойкость криптосистемы, заданной алгоритмом шифрования (4) с учетом условий (a)–(e), накладываемых на множество $\mathcal{E}$, определяется при прямой атаке (т.е. атаке, основанной на декодировании ошибок из $\mathcal{E}$ в коде C) величиной ω. Если ошибки из $\mathcal{E}$ не являются лидерами смежных классов (самыми "легкими" в смежном классе), то их декодирование в коде C (без знания обратного преобразования φ^{-1}) возможно только перебором – либо по словам кода C, либо по множеству ошибок из $\mathcal{E}$. Таким образом,

$$\omega = \min\{2^k, 2^{n-k}, |\mathcal{E}|\}.$$

По построению $|\mathcal{E}| \leqslant 2^{n-k}$. С учетом того, что при декодировании полным перебором по словам кода C сложность декодирования не зависит от $\mathcal{E}$, естественно называть оптимальной криптосистему (4), для которой $|\mathcal{E}| = 2^{n-k}$, и оценивать то, насколько "полно" используются корректирующие свойства кода, величиной

$$\tau = \frac{\log_2 |\mathcal{E}|}{n-k}. \tag{5}$$

Сформулированный нами критерий "полноты" кодовой криптосистемы не является достаточным и даже наиболее важным с точки зрения практического использования криптосистемы. Он не учитывает размеры открытого ключа криптосистемы, которые обычно обсуждались как главный недостаток кодовых криптосистем, – размеры открытого ключа системы, т.е. используемого кода. Поэтому наряду с параметром при оценке криптосистемы мы будем оценивать и размеры применяемого

кода. Далее, кроме атаки на основе полного перебора существует целый ряд неперебо́рных атак, сложность которых приходится учитывать при оценке криптосистем. Кроме того, малое значение параметра τ свидетельствует о том, что "ресурсы" кода, лежащего в основе криптосистемы, используются не в полной мере, а значит, потенциально возможно ее улучшение за счет расширения множества вносимых ошибок, которые впоследствии будут исправляться декодером. И напротив, близкие к единице значения τ позволяют утверждать, что лежащий в основе криптосистемы код используется достаточно эффективно, а значит, дальнейшее расширение множества $\mathcal{E}_0$ за счет внесения в него "тяжелых" векторов ошибок позволит лишь незначительно уменьшить длину ключа.

Например, для классической криптосистемы Мак-Элиса на основе $(1024, 524)$-кода Гоппы τ_{ME} будет оцениваться как

$$\tau_{\mathrm{ME}} \geqslant \frac{\log_2 \binom{1024}{50}}{500} \approx 0{,}5681,$$

а длина открытого ключа при этом $1024 \cdot 524 = 536576$ бит.

§ 5. Криптосистема, основанная на двоичном образе обобщенного кода Рида – Соломона

5.1. Обобщенные коды Рида – Соломона и их двоичные образы. Здесь и далее будем предполагать, что рассматриваемые коды заданы над полем $\mathbb{F}_q$, $q = 2^m$, $m > 0$.

Задача построения множества $\mathcal{E}$, удовлетворяющего условиям (а)–(е) из § 4, для произвольного кода C является достаточно сложной задачей. Тем не менее, двоичный образ обобщенного кода Рида – Соломона (РС-кода), заданного над полем $\mathbb{F}_q$, имеет полиномиальный алгоритм V построения таких множеств достаточно большой мощности.

Вначале напомним определение обобщенного РС-кода $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$.

Определение 1. Пусть задано конечное поле $\mathbb{F}_q$. Выберем ненулевые элементы $v_1, \ldots, v_n \in \mathbb{F}_q$ и различные элементы $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q$. Пусть $\boldsymbol{v} = (v_1, \ldots, v_n)$ и $\alpha = (\alpha_1, \ldots, \alpha_n)$. Для любого $0 \leqslant k \leqslant n$ определим обобщенный код Рида – Соломона как

$$GRS_{n,k}(\alpha, \boldsymbol{v}) = \big\{(v_1 f(\alpha_1), v_2 f(\alpha_2), \ldots, v_n f(\alpha_n)) \mid f(x) \in F_k[x]\big\},$$

где под $F_k[x]$ подразумевается множество полиномов $f(x)$ над полем $\mathbb{F}_q$, степени которых не превосходят $k - 1$.

Известно, что наряду с обычным РС-кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ также является кодом с максимально достижимым расстоянием (МДР-кодом), т.е. имеет минимальное расстояние $d = n - k + 1$. Основная причина, по которой в данной статье рассматривается именно обобщенный РС-код, заключается в том, что для заданных n и k мощность множества различных кодов $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ существенно превосходит количество различных кодов Рида – Соломона, что препятствует структурной атаке на криптосистему, в основе которой лежат $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$. Порождающую матрицу кода $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ будем обозначать через $\boldsymbol{G}'$. Данная матрица задана над полем $\mathbb{F}_q$ и имеет размер $k \times n$.

Зафиксируем некоторый базис $\mathbb{F}_q/\mathbb{F}_2$. Рассмотрим двоичное представление кода $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$, т.е. код, слова которого получаются из слов кода $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ в результате замены символов поля $\mathbb{F}_q$ их двоичным представлением. В результате мы получим двоичный (nm, km)-код с порождающей матрицей $\boldsymbol{G}'_b$ размера $km \times nm$. Обозначим данный код через C_b.

Код C_b в двоичной метрике Хэмминга имеет минимальное расстояние, не меньшее чем у $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$, и при этом способен исправлять любые пакеты ошибок при условии, что данные пакеты ошибок покрывают не более чем t символов принятого слова, если рассматривать их как элементы поля $\mathbb{F}_q$. Если ограничивать только длину пакетов ошибок величинами $1 \leqslant \ell_i \leqslant m$, но при этом снять ограничения на позиции начала и конца пакетов, то число гарантированно исправимых пакетов ошибок будет $\left\lfloor \dfrac{n-k}{4} \right\rfloor = \left\lfloor \dfrac{t}{2} \right\rfloor$. Это следует из того, что никакие $\left\lfloor \dfrac{t}{2} \right\rfloor$ пакетов ошибок длин $1 \leqslant \ell_i \leqslant m$ не исказят более чем t символов поля $\mathbb{F}_q$, а значит, принятый вектор будет исправлен кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ при обратном преобразовании $\mathbb{F}_2^{mn} \mapsto \mathbb{F}_q^n$.

Прежде чем приступать к описанию криптосистемы, введем понятие синхронного и несинхронного пакетов ошибок.

Определение 2. Пакет ошибок длины $1 \leqslant \ell_i \leqslant m$ будем называть синхронным, если для позиции его начала i найдется такое $r \in \mathbb{N} \cup 0$, что одновременно $i \geqslant mr + 1$ и $i + \ell_i - 1 \leqslant m(r+1)$, т.е. все ненулевые элементы пакета попадают в один и только один подвектор $\boldsymbol{e}_{r+1}$ вектора $\boldsymbol{e} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$. В противном случае пакет ошибок будем называть несинхронным.

5.2. Базовое описание протокола криптосистемы. Теперь представим описание криптосистемы с открытым ключом, основанной на двоичном образе обобщенного кода Рида – Соломона. На самом деле, в данном разделе будет представлено высокоуровневое описание предложенной криптосистемы, а представление ее отдельных компонент будет дано в последующих разделах статьи.

Публичная порождающая матрица криптосистемы имеет вид

$$\boldsymbol{G} = \boldsymbol{S}\boldsymbol{G}_b'\boldsymbol{Q}, \tag{6}$$

где $\boldsymbol{G}_b'$ – секретная двоичная порождающая матрица кода C_b, а $\boldsymbol{S}$ – произвольная невырожденная двоичная $(mk \times mk)$-матрица. Двоичная матрица $\boldsymbol{Q}$ размера $mn \times mn$ выбирается согласно теореме 1 из семейства матриц, описание которых представлено в п. 5.3.

Теперь опишем процедуры генерации ключей, шифрования и дешифрования.

- Генерация секретного и публичного ключей:
  1. Выбирается порождающая матрица $\boldsymbol{G}'$ кода $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ и строится ее двоичный образ $\boldsymbol{G}_b'$;
  2. Строится случайная невырожденная двоичная матрица $\boldsymbol{S}$ размера $mk \times mk$;
  3. В соответствии с теоремой 1 строится матрица $\boldsymbol{Q}$ размера $mn \times mn$ и выбирается соответствующий класс вносимых векторов ошибок $\mathcal{V}_i$;
  4. Вычисляется публичная порождающая матрица $\boldsymbol{G} = \boldsymbol{S}\boldsymbol{G}_b'\boldsymbol{Q}$;
  5. Публичным ключом криптосистемы является $(\boldsymbol{G}, \mathcal{V}_i)$;
  6. Секретным ключом криптосистемы является набор $(\boldsymbol{Q}, \boldsymbol{G}_b', \boldsymbol{S})$.
- Шифрование открытого текста $\boldsymbol{x} \in \mathbb{F}_2^{km}$ осуществляется следующим образом:
  1. Выбирается случайный вектор $\boldsymbol{e} \in \mathcal{V}_i \subset \mathbb{F}_2^{mn}$, согласованный с матрицей $\boldsymbol{Q}$, так что вектор ошибок $\boldsymbol{e}\boldsymbol{Q}^{-1}$ исправим кодом с порождающей матрицей $\boldsymbol{G}_b'$;
  2. Вычисляется зашифрованное сообщение $\boldsymbol{y} \in \mathbb{F}_2^{mn}$:

     $$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{G} + \boldsymbol{e}.$$

- Дешифрование вектора $\boldsymbol{y} \in \mathbb{F}_2^{mn}$ осуществляется следующим образом:
  1. Производится умножение $\boldsymbol{y}$ на $\boldsymbol{Q}^{-1}$:

     $$\boldsymbol{y}\boldsymbol{Q}^{-1} = \boldsymbol{x}\boldsymbol{S}\boldsymbol{G}_b' + \boldsymbol{e}\boldsymbol{Q}^{-1};$$

2. Вектор $\boldsymbol{yQ}^{-1}$ преобразуется в q-ичный вектор и декодируется декодером кода $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$, исправляющим t ошибок, откуда находится $\boldsymbol{x}' = \boldsymbol{xS} \in \mathbb{F}_q^k$ – вектор длины k над полем $\mathbb{F}_q$;
3. Вектор $\boldsymbol{x}' \in \mathbb{F}_q^k$ отображается в двоичный вектор $\boldsymbol{x}''$;
4. Зашифрованный текст $\boldsymbol{x}$ находится как

$$\boldsymbol{x} = \boldsymbol{x}''\boldsymbol{S}^{-1}.$$

Как уже было отмечено выше, ключевое требование, которым должна удовлетворять пара $(\boldsymbol{Q}, \boldsymbol{e})$, заключается в том, что вектор $\boldsymbol{eQ}^{-1}$ должен быть исправим кодом с порождающей матрицей $\boldsymbol{G}_b'$, т.е. содержать не более чем t синхронных или $\left\lfloor \frac{t}{2} \right\rfloor$ несинхронных пакетов ошибок длины до m. Далее будет показано, каким образом согласуются структуры векторов $\boldsymbol{e}$ и матриц $\boldsymbol{Q}$ так, чтобы $\boldsymbol{eQ}^{-1} \in \mathcal{E}_0$, где $\mathcal{E}_0$ – множество исправимых кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ ошибок.

5.3. Выбор пары $(\boldsymbol{Q}, \boldsymbol{e})$ в предложенной криптосистеме. Покажем, каким образом нужно выбирать пару $(\boldsymbol{Q}, \boldsymbol{e})$ так, чтобы вектор ошибок $\boldsymbol{eQ}^{-1}$ был исправим двоичным образом обобщенного кода Рида – Соломона (q-ичное представление вектора $\boldsymbol{eQ}^{-1}$ было исправимо кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$) – в этом случае будем говорить, что *вектор $\boldsymbol{e}$ согласован с матрицей $\boldsymbol{Q}$*.

Введем следующее обозначение: символ $\mathcal{A}(\boldsymbol{Q}, \boldsymbol{e})$ означает, что вектор $\boldsymbol{e}$ согласован с матрицей $\boldsymbol{Q}$, т.е. $\boldsymbol{eQ}^{-1}$ содержит не более чем $\left\lfloor \frac{t}{2} \right\rfloor$ несинхронных пакетов ошибок длины до m.

Для простоты введем также следующие обозначения для различных семейств матриц $\boldsymbol{Q}$ и векторов $\boldsymbol{e}$.

Семейства матриц $\boldsymbol{Q}$:

- Будем считать, что матрица $\boldsymbol{Q}$ принадлежит семейству $\mathcal{Q}_1$, если $\boldsymbol{Q} = \operatorname{diag}(\boldsymbol{M})$ – двоичная матрица размера $mn \times mn$, где под $\operatorname{diag}(\boldsymbol{M})$ подразумевается блочно-диагональная $(mn \times mn)$-матрица, на главной диагонали которой стоят невырожденные нижнетреугольные матрицы $\boldsymbol{M}_i$ размеров $m_i \times m_i$, $m+1 \leqslant m_i \leqslant 2m+2$, $\sum m_i = mn$.
- Будем считать, что матрица $\boldsymbol{Q}$ принадлежит семейству $\mathcal{Q}_2$, если $\boldsymbol{Q} = \operatorname{diag}(\boldsymbol{M})$ – двоичная матрица размера $mn \times mn$, где под $\operatorname{diag}(\boldsymbol{M})$ подразумевается блочно-диагональная $(mn \times mn)$-матрица, на главной диагонали которой стоят невырожденные матрицы $\boldsymbol{M}_i$, причем для любых двух соседних матриц $\boldsymbol{M}_{i_1}$ и $\boldsymbol{M}_{i_2}$, стоящих на главной диагонали

  $$\begin{pmatrix} \boldsymbol{M}_{i_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_{i_2} \end{pmatrix}$$

  и имеющих размеры $m_{i_1} \times m_{i_1}$ и $m_{i_2} \times m_{i_2}$, выполняется следующее:
  - $m_{i_1} + m_{i_2} = 2m$;
  - В матрице $\boldsymbol{Q}$ матрицы размеров $m_{i_1} \times m_{i_1}$ и $m_{i_2} \times m_{i_2}$ чередуются;
  - Пусть $\boldsymbol{M}_1$ имеет размер $m_1 \times m_1$, а $\boldsymbol{M}_2$ – размер $m_2 \times m_2$, тогда если $m_1 < m_2$, то в каждом блоке из двух подряд идущих матриц

    $$\begin{pmatrix} \boldsymbol{M}_{i_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_{i_2} \end{pmatrix}$$

    матрицы большего размера являются верхнетреугольными. Если $m_1 > m_2$, то матрицы большего размера являются нижнетреугольными.

Заметим, что ключевое различие между матрицами $\boldsymbol{Q}$ из семейств $\mathcal{Q}_1$ и $\mathcal{Q}_2$ заключается в том, что на матрицы из $\mathcal{Q}_1$ накладываются ограничения на структуру блоков $\boldsymbol{M}_i$ (они должны быть нижнетреугольными), в то время как выбор размеров m_i каждого из блоков остается достаточно гибким: $m+1 \leqslant m_i \leqslant 2m+2$, $\sum m_i = mn$. На элементы $\mathcal{Q}_2$ накладываются ограничения как на размеры соседних матриц $m_{i_1} + m_{i_2} = 2m$, так и на структуру больших матриц $\boldsymbol{M}_i$. Далее будет показано, что накладываемые ограничения на структуру матриц из семейства $\mathcal{Q}_2$ позволяют на этапе шифрования вносить ошибки большего веса, нежели при шифровании с использованием матриц из семейства $\mathcal{Q}_1$.

Семейства векторов $\boldsymbol{e}$:

- Будем считать, что вектор $\boldsymbol{e}$ принадлежит семейству $\mathcal{V}_1$, если $\boldsymbol{e}$ содержит до $\left\lfloor \dfrac{t}{4} \right\rfloor$ несинхронных пакетов ошибок длины до m;
- Будем считать, что вектор $\boldsymbol{e}$ принадлежит семейству $\mathcal{V}_2$, если $\boldsymbol{e}$ содержит до $\left\lfloor \dfrac{t}{3} \right\rfloor$ несинхронных пакетов ошибок длины до m;
- Будем считать, что вектор $\boldsymbol{e}$ принадлежит семейству $\mathcal{V}_3$, если $\boldsymbol{e}$ содержит до $\left\lfloor \dfrac{t}{2} \right\rfloor$ несинхронных пакетов ошибок длины до m.

Согласование $\boldsymbol{e}$ с $\boldsymbol{Q} \in \boldsymbol{\mathcal{Q}_1}$. Ключевым этапом при проектировании криптосистемы, приведенной в п. 5.2, является выбор матрицы $\boldsymbol{Q}$, являющейся составной частью публичного и секретного ключей, а также вектора ошибки $\boldsymbol{e}$, который вносится на этапе шифрования. Для того чтобы согласовать между собой введенные семейства матриц $\mathcal{Q}_1$ и $\mathcal{Q}_2$ с типами векторов ошибки $\mathcal{V}_1$, $\mathcal{V}_2$ и $\mathcal{V}_3$, докажем ряд лемм.

Лемма 1. *Пусть вектор $\boldsymbol{e}$ представляет собой несинхронный m-пакет. Пусть $\boldsymbol{e}' = \boldsymbol{e}\boldsymbol{Q}^{-1}$, где матрица $\boldsymbol{Q} \in \mathcal{Q}_1$. Тогда $\boldsymbol{e}'$ содержит не более четырех синхронных m-пакетов.*

Доказательство. Рассмотрим наихудший случай. Зададим двоичный вектор $\boldsymbol{e}$ длины mn, такой что этот вектор содержит $mn - m$ нулей, а пакет ошибок имеет координату начала, кратную $m-1$. Пусть для простоты данный пакет состоит из единиц. Тогда, если представить $\boldsymbol{e}$ в виде $\boldsymbol{e} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$, $\boldsymbol{e}_i = (e_{i_1}, \ldots, e_{i_m})$, $e_{i_j} \in \mathbb{F}_2$, то $\boldsymbol{e}$ содержит два последовательных вектора $\boldsymbol{e}_i$, $\boldsymbol{e}_{i+1}$, таких что $\boldsymbol{e}_i = (0, 0, \ldots, 0, 1)$, $\boldsymbol{e}_{i+1} = (1, 1, \ldots, 1, 0)$, причем $\mathrm{wt}(\boldsymbol{e}_{i+1}) = m-1$. Все прочие $\boldsymbol{e}_j$, $j \notin \{i, i+1\}$, являются нулевыми векторами длины m. Пусть в матрице $\boldsymbol{Q}^{-1}$, соответствующей ненулевому участку вектора $\boldsymbol{e}$, находятся две матрицы $\boldsymbol{M}_{i_1}$, $\boldsymbol{M}_{i_2}$ размеров $m_{i_1} \times m_{i_1}$ и $m_{i_2} \times m_{i_2}$ соответственно. Рассмотрим векторы $\boldsymbol{e}'_i = (0, 0, \ldots 0, \boldsymbol{e}_i)$, $\boldsymbol{e}'_{i+1} = (\boldsymbol{e}_{i+1}, 0, \ldots, 0)$ длин m_{i_1} и m_{i_2} соответственно. При вычислении $\boldsymbol{e}' = \boldsymbol{e}\boldsymbol{Q}^{-1}$ участок вектора $\boldsymbol{e}'$, соответствующий произведению $(\boldsymbol{e}'_i, \boldsymbol{e}'_{i+1})$ на $\boldsymbol{Q}^{-1}$, будет иметь вид

$$(\widehat{\boldsymbol{e}}_i, \widehat{\boldsymbol{e}}_{i+1}) = (\boldsymbol{e}'_i \boldsymbol{M}_{i_1}, \boldsymbol{e}'_{i+1} \boldsymbol{M}_{i_2}).$$

Так как вектор $\widehat{\boldsymbol{e}}_i$ содержит единицу на последней позиции, то $\boldsymbol{e}'_i \boldsymbol{M}_{i_1}$ совпадает с последней строкой матрицы $\boldsymbol{M}_{i_1}$, имеющей вес до m_{i_1}. В худшем случае (с точки зрения распространения пакетов ошибок) вектор $\widehat{\boldsymbol{e}}_i$ начинается с 1. Вектор $\widehat{\boldsymbol{e}}_{i+1}$ представляет собой произведение вектора, содержащего первые $m-1$ единицы, а остальные $m_{i_2} - m + 1$ его символов равны 0. Таким образом, $\boldsymbol{e}'_{i+1} \boldsymbol{M}_{i_2}$ имеет вид

$$\widehat{\boldsymbol{e}}_{i+1} = (\widehat{e}_{i+1,1}, \widehat{e}_{i+1,2}, \ldots, \widehat{e}_{i+1,m-1}, 0, \ldots, 0),$$

где $\widehat{e}_{i+1,m-1}$ могут быть ненулевыми. В худшем случае $\widehat{e}_{i+1,m-1} = 1$.

Таким образом, вектор $(\widehat{\boldsymbol{e}}_i, \widehat{\boldsymbol{e}}_{i+1})$ длины $m_{i_1} + m_{i_2}$, где $2m+2 \leqslant m_{i_1} + m_{i_2} \leqslant 4m+4$, содержит пакет ошибок длины до $m_{i_1} + m - 1 \leqslant 3m+1$. Очевидно, что данный пакет ошибок покрывается максимум четырьмя синхронными m-пакетами, а значит, преобразование $\boldsymbol{Q}^{-1}$ не более чем в 4 раза увеличивает вес вектора ошибки

(в q-ичной метрике Хэмминга), который впоследствии должен быть декодирован кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$.

Если участку вектора $\boldsymbol{e}$, на котором расположены $(\boldsymbol{e}_i, \boldsymbol{e}_{i+1})$, в матрице $\boldsymbol{Q}^{-1}$ соответствует единственная матрица $\boldsymbol{M}_i$ размера не более чем $2m+2$, то вектор $(\boldsymbol{e}'_i, \boldsymbol{e}'_{i+1})\boldsymbol{M}_i$ покрывает не более чем четыре символа поля $\mathbb{F}_q$.

По построению матрицы $\boldsymbol{Q}$ никакому пакету ошибок длины не более m на участке $(\boldsymbol{e}_i, \boldsymbol{e}_{i+1})$ длины $2m$ вектора $\boldsymbol{e}$ не может соответствовать более двух блочных подматриц $\boldsymbol{M}_{i_1}$, $\boldsymbol{M}_{i_2}$.

Таким образом, никакой пакет ошибок длины m при преобразовании $\boldsymbol{Q}^{-1}$ не покрывает более чем четыре символа поля $\mathbb{F}_q$. ▲

Таким образом, если $\boldsymbol{Q} \in \mathcal{Q}_1$ и $\boldsymbol{e} \in \mathcal{V}_1$, то выполнено $\mathcal{A}(\boldsymbol{Q}, \boldsymbol{e})$.

Покажем теперь, какие ограничения сверху на m_i в матрицах $\boldsymbol{Q} \in \mathcal{Q}_1$ необходимо накладывать, чтобы произвольный пакет ошибок длины до m при преобразовании $\boldsymbol{Q}^{-1}$ покрывал как можно меньше символов поля $\mathbb{F}_q$, что позволило бы увеличить число вносимых ошибок. Ограничение снизу $m_i \geqslant m+1$ будет сохранено для гарантии того, что никакому пакету ошибок длины не более m на участке $(\boldsymbol{e}_i, \boldsymbol{e}_{i+1})$ длины $2m$ не может соответствовать более двух блочных матриц $\boldsymbol{M}_{i_1}$, $\boldsymbol{M}_{i_2}$ в $\boldsymbol{Q}^{-1}$. Напомним, что в худшем случае умножение пакета ошибок на матрицу $\boldsymbol{Q}^{-1}$ формирует пакет длины $m_i + m - 1 \geqslant 2m$. Ясно, что таким пакетом можно покрыть не более чем три последовательных символа поля $\mathbb{F}_q$. При длине пакета ошибок не более чем $2m+1$, число последовательно покрытых символов поля векторов длины m в $\boldsymbol{e}\boldsymbol{Q}^{-1}$ не превышает трех. Таким образом, если получить ограничение сверху на m_i из

$$m_i + m - 1 \leqslant 2m + 1,$$

т.е. $m_i \leqslant m+2$, то вместо внесения $\left\lfloor \frac{t}{4} \right\rfloor$ пакетов ошибок длины до m в вектор $\boldsymbol{e}$ можно вносить $\left\lfloor \frac{t}{3} \right\rfloor$ пакетов ошибок длины до m. Отменим, что внесение $\left\lfloor \frac{t}{2} \right\rfloor$ ошибок уже не гарантирует декодируемость вектора $\boldsymbol{e}\boldsymbol{Q}^{-1}$ при описанной ранее структуре матрицы $\boldsymbol{Q}$.

Таким образом, справедлива

Лемма 2. *Пусть вектор $\boldsymbol{e}$ представляет собой несинхронный m-пакет. Пусть $\boldsymbol{e}' = \boldsymbol{e}\boldsymbol{Q}^{-1}$, где матрица $\boldsymbol{Q} \in \mathcal{Q}_1$, и при этом для размеров m_i блоков $\boldsymbol{M}_i$ справедливо неравенство $m+1 \leqslant m_i \leqslant m+2$. Тогда $\boldsymbol{e}'$ содержит не более трех синхронных m-пакетов.*

Таким образом, если $m+1 \leqslant m_i \leqslant m+2$, $\boldsymbol{Q} \in \mathcal{Q}_1$ и $\boldsymbol{e} \in \mathcal{V}_2$, то $\mathcal{A}(\boldsymbol{Q}, \boldsymbol{e})$.

В общем случае при внесении не более чем $\left\lfloor \frac{t}{\ell} \right\rfloor$, $\ell \geqslant 2$, пакетов ошибок длины до m для декодируемости вектора $\boldsymbol{e}\boldsymbol{Q}^{-1}$ (т.е. согласованности вектора $\boldsymbol{e}$ с матрицей $\boldsymbol{Q} \in \mathcal{Q}_1$) при ограничении снизу $m_i \geqslant m+1$ следует ограничение сверху

$$m_i \leqslant (\ell - 1)m - m + 2.$$

Согласование $\boldsymbol{e}$ с $\boldsymbol{Q} \in \mathcal{Q}_2$. Ранее было показано, что размеры m_i блоков $\boldsymbol{M}_i$ матрицы $\boldsymbol{Q} \in \mathcal{Q}_1$ существенно влияют на число пакетов ошибок, которые можно вносить при шифровании. Ясно также, что при отсутствии ограничений на индексы начала пакетов ошибок можно исправлять до $\left\lfloor \frac{t}{2} \right\rfloor$ пакетов ошибок длины до m, где t – число ошибок, исправимых кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$. Однако ранее было показано, что при единственном ограничении $m_i \geqslant m+1$, где m_i – размеры квадратных матриц, входящих в состав $\boldsymbol{Q} \in \mathcal{Q}_1$, наибольшее число пакетов ошибок, вносимых на этапе шифрования, не может превышать $\left\lfloor \frac{t}{3} \right\rfloor$. Только в этом случае можно га-

рантировать возможность их исправления кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$ после применения преобразования $\boldsymbol{Q}^{-1}$.

Покажем, что если матрица $\boldsymbol{Q} \in \mathcal{Q}_2$ и при этом для размеров m_{i_1} и m_{i_2} любых двух соседних невырожденных матриц $\boldsymbol{M}_{i_1}$ и $\boldsymbol{M}_{i_2}$ выполняется $m_{i_1} + m_{i_2} = 2m$, то с учетом ограничений на большие матрицы из семейства $\mathcal{Q}_2$ матрица $\boldsymbol{Q}$ согласована с $\boldsymbol{e} \in \mathcal{V}_3$, т.е. на этапе шифрования допустимым было бы внесение максимального числа $\left\lfloor \dfrac{t}{2} \right\rfloor$ несинхронных пакетов ошибок.

Лемма 3. *Если матрица $\boldsymbol{Q} \in \mathcal{Q}_2$, а вектор $\boldsymbol{e}$ представляет собой несинхронный m-пакет, то вектор $\boldsymbol{e}\boldsymbol{Q}^{-1}$ содержит не более двух синхронных m-пакетов.*

Доказательство. Очевидно, что для того чтобы $\boldsymbol{e}\boldsymbol{Q}^{-1}$ содержал не более двух синхронных m-пакетов, необходимо и достаточно, чтобы умножение вектора $(\boldsymbol{e}_i, \boldsymbol{e}_{i+1})$ длины $2m$, содержащего на произвольных m подряд идущих позициях пакет ошибок длины до m, на соответствующий участок длины $2m$ матрицы $\boldsymbol{Q}^{-1}$ не приводило к "размножению" пакетов.

Это, очевидно, достигается в том случае, когда соответствующий участок матрицы $\boldsymbol{Q}^{-1}$ имеет вид

$$\begin{pmatrix} \boldsymbol{M}_{i_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_{i_2} \end{pmatrix},$$

где $\boldsymbol{M}_{i_1}$ и $\boldsymbol{M}_{i_2}$ – квадратные матрицы размеров $m_{i_1} \times m_{i_1}$ и $m_{i_2} \times m_{i_2}$, и $m_{i_1} + m_{i_2} = 2m$. Если при этом в $\boldsymbol{Q}$ матрицы размеров $m_{i_1} \times m_{i_1}$ и $m_{i_2} \times m_{i_2}$ чередуются, то никакой пакет ошибок веса m не будет стоять на пересечении более чем двух матриц в $\boldsymbol{Q}$.

Если, кроме того, выполняются дополнительные ограничения на структуру больших матриц $\boldsymbol{M}_i$ (они являются верхнетреугольными, если размер первой блочной матрицы $\boldsymbol{M}_1$ меньше размера $\boldsymbol{M}_2$, и нижнетреугольными, если размер первой блочной матрицы $\boldsymbol{M}_1$ больше размера $\boldsymbol{M}_2$), то каков бы ни был пакет ошибок, лежащий в $(\boldsymbol{e}_i, \boldsymbol{e}_{i+1})$, при умножении на $\boldsymbol{Q}^{-1}$ он не "размножится" на соседние символы, а потому вектор $\boldsymbol{e}' = \boldsymbol{e}\boldsymbol{Q}^{-1}$ будет иметь ту же структуру, что и вектор $\boldsymbol{e}$, который генерируется на этапе шифрования.

Единственное отличие вектора $\boldsymbol{e}'$ от $\boldsymbol{e}$ будет заключаться в том, что длины пакетов ошибок в $\boldsymbol{e}'$ могут достигать $2m$, однако вес Хэмминга вектора $\boldsymbol{e}'$, вычисленный над полем $\mathbb{F}_q$, не будет превышать t, что гарантирует его исправимость кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$. ▲

Согласование $\boldsymbol{e}$ с $\boldsymbol{Q}$ – основной результат. Объединяя леммы 1–3, сформулируем теорему, связывающую между собой семейства $\mathcal{Q}_1$ и $\mathcal{Q}_2$ с семействами $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$ векторов $\boldsymbol{e}$ так, чтобы $\boldsymbol{e}\boldsymbol{Q}^{-1}$ содержал не более чем t синхронных пакетов ошибок длины до m, т.е. был исправим кодом $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$.

Теорема 1. *Справедливы следующие утверждения*:

- *Если $\boldsymbol{Q} \in \mathcal{Q}_1$, $\boldsymbol{e} \in \mathcal{V}_1$ и для всех блоков $\boldsymbol{M}_i$ размеров $m_i \times m_i$ выполняется $m + 1 \leqslant m_i \leqslant 2m + 2$, $\sum m_i = mn$, то имеет место $\mathcal{A}(\boldsymbol{Q}, \boldsymbol{e})$;*
- *Если $\boldsymbol{Q} \in \mathcal{Q}_1$, $\boldsymbol{e} \in \mathcal{V}_1 \cup \mathcal{V}_2$ и для всех блоков $\boldsymbol{M}_i$ размеров $m_i \times m_i$ выполняется $m + 1 \leqslant m_i \leqslant m + 2$, $\sum m_i = mn$, то имеет место $\mathcal{A}(\boldsymbol{Q}, \boldsymbol{e})$;*
- *Если $\boldsymbol{Q} \in \mathcal{Q}_2$ и $\boldsymbol{e} \in \mathcal{V}_1 \cup \mathcal{V}_2 \cup \mathcal{V}_3$, то имеет место $\mathcal{A}(\boldsymbol{Q}, \boldsymbol{e})$.*

Таким образом, на этапе проектирования криптографической системы, основанной на двоичном образе обобщенного кода Рида – Соломона, представленной в п. 5.2, разработчик выбирает согласованную в соответствии с теоремой 1 пару $(\boldsymbol{Q}, \boldsymbol{e})$. Выбор пары позволяет регулировать гибкость параметров, определяющих матрицу $\boldsymbol{Q}$, и число вносимых на этапе шифрования ошибок.

Следует особо отметить, что в отличие от классической криптосистемы Мак-Элиса, где публичная порождающая матрица задает линейный код, эквивалентный секретному, в нашем случае это не так: умножение порождающей матрицы $\boldsymbol{G}'_b$ на $\boldsymbol{Q} \in \mathcal{Q}_1 \cup \mathcal{Q}_2$ справа задает преобразование столбцов в $\boldsymbol{G}'_b$, а потому матрица $\boldsymbol{S}\boldsymbol{G}'_b\boldsymbol{Q}$ не является порождающей матрицей двоичного образа кода $GRS_{n,k}(\boldsymbol{\alpha}, \boldsymbol{v})$. Эквивалентность кодов сохранилась бы в том случае, если бы $\boldsymbol{Q}$ являлась блочной перестановкой длины n, а длина блока была равна m, т.е. задавала перестановку символов поля $\mathbb{F}_q$. Исходя из того, что код C_b исправляет максимальное число пакетов ошибок длины до m, с большой вероятностью данное множество пакетов будет не исправимо кодом с порождающей матрицей $\boldsymbol{S}\boldsymbol{G}'_b\boldsymbol{Q}$, что является ключевым фактором, на котором основана предложенная криптосистема.

Далее будут рассмотрены атаки декодирования на предложенный класс криптосистем.

5.4. Анализ некоторых атак. При рассмотрении атак мы будем отталкиваться от того, что на этапе шифрования в вектор $\boldsymbol{e}$ вносится до $\left\lfloor \frac{t}{4} \right\rfloor$ пакетов ошибок длины до m, хотя все полученные результаты легко обобщаются для произвольного числа пакетов $\left\lfloor \frac{t}{\ell} \right\rfloor$, $\ell \geqslant 2$.

Прямые атаки. Напомним, что прямые атаки сводятся к перебору либо информационных векторов $\boldsymbol{x}$, либо векторов ошибок $\boldsymbol{e}$. Сложность прямых атак можно оценить сверху следующим образом:

- Максимальное число раундов для восстановления $\boldsymbol{x}$: $\min\{2^{mk}, 2^{m(n-k)}\}$, на каждом раунде происходит умножение вектора длины mk или $m(n-k)$ на публичную проверочную или порождающую матрицу, требующее $m^2k(n-k)$ операций;
- Максимальное число раундов для восстановления $\boldsymbol{e}$: $\binom{mn}{\lfloor t/4 \rfloor} 2^{m-1}$, на каждом раунде вектор $\boldsymbol{e}$ вычитается из принятого вектора $\boldsymbol{y}$ и считается синдром, это требует $m^2k(n-k)$ операций.

Таким образом, сложность C_{dir} прямой атаки можно оценить как

$$C_{\text{dir}} = \mathcal{O}\left(m^2k(n-k) \cdot \min\left\{ 2^{mk},\ 2^{m(n-k)},\ \binom{mn}{\left\lfloor \frac{t}{4} \right\rfloor} 2^{m-1} \right\} \right).$$

Атака на основе декодирования по информационным совокупностям. Как известно, для классической криптосистемы Мак-Элиса атака декодирования по информационным совокупностям является наиболее эффективной – именно она определяет сложность раскрытия криптосистемы и влияет на выбор параметров кодов (а значит, и длины публичного и секретного ключей), необходимых для достижения заданного уровня стойкости.

Поскольку в вектор $\boldsymbol{e}$ длины mn вносится $\left\lfloor \frac{t}{4} \right\rfloor$ пакетов ошибок длины до m, то для поиска информационной совокупности, свободной от ошибок, необходимо найти $\left\lfloor \frac{t}{4} \right\rfloor$ индексов начала каждого из пакетов ошибок и считать, что длина каждого пакета ошибок равна m. Количество раундов для нахождения данных позиций не превосходит $\binom{mn}{\lfloor t/4 \rfloor}$. Таким образом, сложность нахождения информационной совокупности, свободной от ошибок, есть

$$C_{\text{ISD}} = \binom{mn}{\left\lfloor \frac{t}{4} \right\rfloor} m^2k(n-k).$$

Если искать информационную совокупность среди дополнения к множеству из $\left\lfloor \frac{t}{4} \right\rfloor$ непересекающихся пакетов ошибок длины m, то последнюю оценку можно уточнить:

$$C_{\mathrm{ISD}} = \\ = \frac{m(n-1)(m(n-1)-m)(m(n-1)-2m)\dots\left(m(n-1)-m\left\lfloor \frac{t}{4} \right\rfloor\right)}{\left(\left\lfloor \frac{t}{4} \right\rfloor\right)!} m^2 k(n-k).$$

Синдромная атака. Суть синдромной атаки состоит в том, чтобы по публичной порождающей матрице $\boldsymbol{G}$ вычислить публичную проверочную матрицу $\boldsymbol{H}$, а затем свести задачу нахождения $\boldsymbol{x}$ из соотношения $\boldsymbol{y} = \boldsymbol{x}\boldsymbol{G} + \boldsymbol{e}$ к решению соответствующего синдромного уравнения за счет умножения обеих частей на $\boldsymbol{H}^{\mathrm{T}}$. Так как

$$\boldsymbol{G} = \boldsymbol{S}\boldsymbol{G}'_b\boldsymbol{Q},$$

то

$$\boldsymbol{H} = \boldsymbol{L}\boldsymbol{H}'_b(\boldsymbol{Q}^{-1})^{\mathrm{T}},$$

где $\boldsymbol{H}'_b$ – проверочная матрица, соответствующая порождающей матрице $\boldsymbol{G}'_b$, а $\boldsymbol{L}$ – некоторая невырожденная матрица размера $m(n-k)\times m(n-k)$ над полем $\mathbb{F}_2$. Ясно, что матрица $\boldsymbol{L}$ не влияет на свойства кода. Поэтому будем полагать, что $\boldsymbol{L} = \boldsymbol{I}$. В этом случае синдром $\boldsymbol{Z}$ зашифрованного текста $\boldsymbol{y}$ имеет вид

$$\boldsymbol{Z} = \boldsymbol{y}\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{e}\boldsymbol{Q}^{-1}(\boldsymbol{H}'_b)^{\mathrm{T}}.$$

Однако ввиду произвольности в выборе $\boldsymbol{Q}$ проверочная матрица $\boldsymbol{H}$ может соответствовать коду с расстоянием значительно меньшим, чем расстояние кода C_b. Таким образом, применение синдромной атаки не позволяет гарантированно найти вектор пакетов ошибок, сгенерированный на этапе шифрования.

5.5. Длина ключей. Сложность C_{comp} криптоанализа предложенной в статье криптосистемы мы будем оценивать сверху величиной

$$C_{\mathrm{comp}} = \mathcal{O}(\min\{C_{\mathrm{dir}}, C_{\mathrm{ISD}}\}).$$

Таким образом, для получения заданной стойкости W системы необходимо выбрать такой (n,k)-код (возможно, укороченный) $GRS_{n,k}(\boldsymbol{\alpha},\boldsymbol{v})$ над полем $\mathbb{F}_q$, чтобы $W \leqslant C_{\mathrm{comp}}$. Длина публичного ключа при этом составит $L_{\mathrm{pub}} = knm^2$. Таким образом, оптимальная с точки зрения длины ключа криптосистема, имеющая стойкость W, определяется тройкой параметров (n,k,m), $n \leqslant 2^m - 1 = q-1$, $k < n$, для которых

$$\begin{cases} knm^2 \to \min, \\ C_{\mathrm{comp}} \geqslant W, \\ n \leqslant 2^m - 1, \\ 0 < k < n. \end{cases}$$

Рассмотрим несколько примеров.

Пример 1. Пусть $W = 2^{72}$. Рассмотрим укороченный обобщенный $(76,18)$-код Рида–Соломона над полем $\mathbb{F}_q$, $q = 2^7$, полученный из обобщенного кода Рида–Соломона над полем $\mathbb{F}_q$. Данный код имеет расстояние 59 и исправляет 29 любых независимых q-ичных ошибок. Рассмотрим далее двоичный образ этого кода, представив каждый элемент поля $\mathbb{F}_q$ в виде двоичного вектора длины 7. В результате получим двоичный $(532,126)$-код C, исправляющий до 29 пакетов ошибок длины

до 7. Если взять данный код в качестве основы для построения описанной выше криптосистемы, то сложность криптоанализа системы оценивается как

1. $2^{mk}m^2k(n-k) > 2^{141}$ – сложность атаки на основе перебора информационных векторов;
2. $2^{m(n-k)}m^2k(n-k) > 2^{421}$ – сложность атаки на основе перебора информационных векторов для двойственного кода;
3. $\binom{mn}{\lfloor t/4 \rfloor}2^{m-1}m^2k(n-k) > 2^{72}$ – сложность атаки на основе перебора всех векторов ошибок;
4. Сложность атаки по информационным совокупностям оценивается как

$$C_{\text{ISD}} = \frac{7 \cdot 75 \cdot (7 \cdot 75 - 7) \cdot (7 \cdot 75 - 14) \cdot \ldots \cdot (7 \cdot 75 - 49)}{7!} \cdot 49 \cdot 18 \cdot 58 > 2^{75}.$$

Таким образом, стойкость криптосистемы $W_c \approx 2^{72} \approx W$. При этом длина ключа составляет $L_{\text{pub}} = 76 \cdot 18 \cdot 7^2 = 67032$ бит, что более чем в 8 раз меньше длины ключа криптосистемы Мак-Элиса, основанной на $(1024, 524, 101)$-коде Гоппы и имеющей стойкость 2^{72}.

"Полнота" множества вносимых ошибок согласно формуле (5) оценивается снизу величиной

$$\tau_{GRS} = \frac{\log_2 |\mathcal{E}|}{m(n-k)} = \frac{\log_2 \left(\binom{mn}{\lfloor t/4 \rfloor} 2^{m-1} \right)}{m(n-k)} \approx 0{,}1405,$$

что значительно уступает оценке данной величины для криптосистемы Мак-Элиса $\tau_{\text{ME}} \approx 0{,}5681$. Это в первую очередь говорит о том, что двоичный образ обобщенного кода Рида – Соломона способен исправлять значительно более широкое множество ошибок, нежели генерируемое в рамках данной криптосистемы. А значит, в перспективе возможно дальнейшее уменьшение длины публичного ключа, что и будет сделано в последующих примерах.

Приведем еще один пример параметров криптосистемы в предположении того, что на этапе шифрования вносится $\left\lfloor \frac{t}{3} \right\rfloor$ пакета ошибок длины до m. Напомним, что при этом размеры матриц $\boldsymbol{M}_i$ выбираются из множества $\{m+1, m+2\}$. Оценки сложности криптоанализа при этом очевидно получаются из аналогичных соотношений для случая внесения $\left\lfloor \frac{t}{4} \right\rfloor$ пакетов ошибок.

Пример 2. Пусть $W = 2^{72}$. Рассмотрим обобщенный $(63, 15)$-код Рида – Соломона над полем $\mathbb{F}_q$, $q = 2^6$, полученный из обобщенного кода Рида – Соломона над полем $\mathbb{F}_q$. Данный код имеет расстояние 48 и исправляет 24 любые q-ичные независимые ошибки. Рассмотрим далее двоичный образ этого кода, представив каждый элемент поля $\mathbb{F}_q$ в виде двоичного вектора длины 6. В результате получим двоичный $(378, 90)$-код C, исправляющий до 24 пакетов ошибок длины до 6. Если взять данный код в качестве основы для построения описанной выше криптосистемы, то сложность криптоанализа системы оценивается как

1. $2^{mk}m^2k(n-k) > 2^{104}$ – сложность атаки на основе перебора информационных векторов;
2. $2^{m(n-k)}m^2k(n-k) > 2^{302}$ – сложность атаки на основе перебора информационных векторов для двойственного кода;
3. $\binom{mn}{\lfloor t/3 \rfloor}2^{m-1}m^2k(n-k) > 2^{72}$ – сложность атаки на основе перебора всех векторов ошибок;

4. Сложность атаки по информационным совокупностям оценивается как

$$C_{\text{ISD}} = \frac{6 \cdot 62 \cdot (6 \cdot 62 - 6) \cdot (6 \cdot 62 - 12) \cdot \ldots \cdot (6 \cdot 62 - 96)}{8!} \cdot 36 \cdot 15 \cdot 48 > 2^{75}.$$

Таким образом, стойкость криптосистемы $W_c \approx 2^{72} \approx W$. При этом длина ключа составляет $L_{\text{pub}} = 63 \cdot 15 \cdot 6^2 = 34020$ бит, что более чем в 15 раз меньше длины ключа криптосистемы Мак-Элиса, основанной на $(1024, 524, 101)$-коде Гоппы и имеющей стойкость 2^{72}.

"Полнота" множества вносимых ошибок для этой криптосистемы согласно формуле (5) оценивается снизу величиной

$$\tau_{GRS} = \frac{\log_2 |\mathcal{E}|}{m(n-k)} = \frac{\log_2 \left(\binom{mn}{\lfloor t/3 \rfloor} 2^{m-1} \right)}{m(n-k)} \approx 0{,}19147,$$

что по-прежнему меньше, чем у криптосистемы Мак-Элиса, но больше, чем у криптосистемы, основанной на укороченном обобщенном $(76, 18)$-коде Рида – Соломона.

В заключение этого параграфа рассмотрим еще один пример параметров криптосистемы в предположении того, что на этапе шифрования вносится $\left\lfloor \frac{t}{2} \right\rfloor$ пакетов ошибок длины до m. Напомним, что при этом накладываются некоторые дополнительные ограничения на матрицу $\boldsymbol{Q}$, которые были рассмотрены ранее. Оценки сложности криптоанализа при этом очевидно получаются из аналогичных соотношений для случая внесения $\left\lfloor \frac{t}{4} \right\rfloor$ или $\left\lfloor \frac{t}{3} \right\rfloor$ пакетов ошибок.

Пример 3. Пусть $W = 2^{72}$. Рассмотрим укороченный обобщенный $(46, 10)$-код Рида – Соломона над полем $\mathbb{F}_q$, $q = 2^6$. Данный код имеет расстояние 37 и исправляет 18 любых q-ичных независимых ошибок. Рассмотрим далее двоичный образ этого кода, представив каждый элемент поля $\mathbb{F}_q$ в виде двоичного вектора длины 6. В результате получим двоичный $(276, 60)$-код C, исправляющий до 18 пакетов ошибок длины до 6. Если взять данный код в качестве основы для построения описанной выше криптосистемы, то сложность криптоанализа системы оценивается как

1. $2^{mk} m^2 k(n-k) > 2^{73}$ – сложность атаки на основе перебора информационных векторов;
2. $2^{m(n-k)} m^2 k(n-k) > 2^{229}$ – сложность атаки на основе перебора информационных векторов для двойственного кода;
3. $\binom{mn}{\lfloor t/2 \rfloor} 2^{m-1} m^2 k(n-k) \approx 2^{73}$ – сложность атаки на основе перебора всех векторов ошибок;
4. Сложность атаки по информационным совокупностям оценивается как

$$C_{\text{ISD}} = \frac{6 \cdot 45 \cdot (6 \cdot 45 - 6) \cdot (6 \cdot 45 - 12) \cdot \ldots \cdot (6 \cdot 45 - 63)}{9!} \cdot 36 \cdot 10 \cdot 36 > 2^{74}.$$

Таким образом, стойкость криптосистемы $W_c \approx 2^{73} > W$. При этом длина ключа составляет $L_{\text{pub}} = 46 \cdot 10 \cdot 6^2 = 16560$ бит, что более чем в 32 раза меньше длины ключа криптосистемы Мак-Элиса, основанной на $(1024, 524, 101)$-коде Гоппы и имеющей стойкость 2^{72}.

"Полнота" множества вносимых ошибок для такой криптосистемы согласно формуле (5) оценивается снизу величиной

$$\tau_{GRS} = \frac{\log_2 |\mathcal{E}|}{m(n-k)} = \frac{\log_2 \left(\binom{mn}{\lfloor t/2 \rfloor} 2^{m-1} \right)}{m(n-k)} \approx 0{,}2746,$$

что по-прежнему меньше, чем у криптосистемы Мак-Элиса, но больше, чем у криптосистем, основанных на укороченных обобщенных $(76, 18)$- и $(63, 15)$-кодах Рида – Соломона.

§ 6. Заключение

В статье рассмотрена общая постановка задачи построения криптосистемы с открытым ключом на основе кодов, исправляющих ошибки.

Сформулированы свойства, которым должна соответствовать кодовая криптосистема для того, чтобы обеспечивать требуемый уровень стойкости.

Предложен критерий τ для сравнения криптосистем между собой, оценивающий взаимосвязь между мощностью множества ошибок, вводимых в криптосистему для обеспечения ее стойкости, и количеством проверочных символов у кода, лежащего в основе криптосистемы. Таким образом, величину $1-\tau$ можно трактовать как меру потенциала улучшения криптосистемы за счет дальнейшего расширения множества вносимых ошибок.

Чтобы продемонстрировать теоретическую возможность построения криптосистем, для которых возможно выполнение сформулированных в статье свойств, описана конструкция, удовлетворяющая предложенному набору условий. Эта конструкция основана на двоичных образах обобщенных кодов Рида – Соломона. Продемонстрировано, что данная конструкция имеет меньшую длину ключа для заданных параметров стойкости по сравнению с криптосистемой Мак-Элиса на основе двоичных кодов Гоппы.

В результате анализа свойств криптосистемы было показано, что наиболее перспективными являются криптосистемы, где количество вносимых ошибок значительно больше, чем половина минимального расстояния, при этом декодирование по информационным совокупностям перестает быть наиболее эффективной стратегией атаки по декодированию.

Результаты статьи показывают, что построение кодовых криптосистем на основе использования маскирующих векторов (векторов ошибки) малого веса Хэмминга не является эффективным, поскольку такие системы чувствительны к атаке на основе поиска свободной от ошибок информационной совокупности. Использование векторов ошибок, не являющихся самыми легкими в своих смежных классах, позволяют существенно снизить эффективность атаки по информационным совокупностям. При этом встает задача поиска преобразований, отображающих легкие представители смежных классов кодов в более тяжелые. Задача эта в теории кодирования, насколько нам известно, не решалась. Мы надеемся, что эта задача может оказаться плодотворной как в криптографии, так и в других приложениях теории помехоустойчивого кодирования.

СПИСОК ЛИТЕРАТУРЫ

1. *McEliece R.J.* A Public-Key Cryptosystem Based on Algebraic Coding Theory // DSN Progress Report 42-44. Jet Propulsion Lab., California Inst. of Technology, Pasadena, CA. 1978. P. 114–116. Available at `https://tmo.jpl.nasa.gov/progress_report2/42-44/44N.PDF`
2. *Kabatianskii G., Krouk E., Smeets B.* A Digital Signature Scheme Based on Random Error-Correcting Codes // Crytography and Coding (Proc. 6th IMA Int. Conf. on Cryptography and Coding. Cirencester, UK. Dec. 17–19, 1997). Lect. Notes Comput. Sci. V. 1355. Berlin: Springer, 1997. P. 161–167. `https://doi.org/10.1007/BFb0024461`
3. *Rivest R.L., Shamir A., Adleman L.* A Method for Obtaining Digital Signatures and Public-Key Cryptosystems // Commun. ACM. 1978. V. 21. № 2. P. 120–126. `https://doi.org/10.1145/359340.359342`

4. *El Gamal T.* A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms // IEEE Trans. Inform. Theory. 1985. V. 31. № 4. P. 469–472. https://doi.org/10.1109/TIT.1985.1057074
5. *Véron P.* Code Based Cryptography and Steganography // Algebraic Informatics (Proc. 5th Int. Conf. on Algebraic Informatics (CAI'2013). Porquerolles, France. Sept. 3–6, 2013). Lect. Notes Comput. Sci. V. 8080. Berlin: Springer, 2013. P. 9–46. https://doi.org/10.1007/978-3-642-40663-8_5
6. *Berger T.P., Cayrel P.L., Gaborit P., Otmani A.* Reducing Key Length of the McEliece Cryptosystem // Progress in Cryptology — AFRICACRYPT 2009 (Proc. 2nd Int. Conf. on Cryptology in Africa. Gammarth, Tunisia. June 21–25, 2009). Lect. Notes Comput. Sci. V. 5580. Berlin: Springer, 2009. P. 77–97. https://doi.org/10.1007/978-3-642-02384-2_6
7. *Faugère J.-C., Otmani A., Perret L., Tillich J.-P.* Algebraic Cryptanalysis of McEliece Variants with Compact Keys // Advances in Cryptology — EUROCRYPT 2010 (Proc. 29th Annu. Int. Conf. on the Theory and Applications of Cryptographic Techniques. French Riviera. May 30 – June 3, 2010). Lect. Notes Comput. Sci. V. 6110. Berlin: Springer, 2010. P. 279–298. https://doi.org/10.1007/978-3-642-13190-5_14
8. *Kocher P.C.* Timing Attacks on Implementations of Diffie–Hellman, RSA, DSS, and Other Systems // Advances in Cryptology — CRYPTO'96 (Proc. 16th Annu. Int. Cryptology Conf. Santa Barbara, CA, USA. Aug. 18–22, 1996). Lect. Notes Comput. Sci. V. 1109. Berlin: Springer, 1996. P. 104–113. https://doi.org/10.1007/3-540-68697-5_9
9. *Barker E.* NIST Special Publication (SP) 800-57 Part 1 Revision 4. Recommendation for Key Management – Part 1: General. National Inst. of Standards and Technology, Gaithersburg, MD, USA, 2016. https://doi.org/10.6028/NIST.SP.800-57pt1r4
10. *Chen L., Jordan S., Liu Y.-K., Moody D., Peralta R., Perlner R., Smith-Tone D.* Report on Post-Quantum Cryptography. NIST Internal Report 8105. National Inst. of Standards and Technology, Gaithersburg, MD, USA, 2016. https://doi.org/10.6028/NIST.IR.8105
11. *Shor P.W.* Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer // SIAM Rev. 1999. V. 41. № 2. P. 303–332. https://doi.org/10.1137/S0036144598347011
12. *Berlekamp E., McEliece R., van Tilborg H.* On the Inherent Intractability of Certain Coding Problems // IEEE Trans. Inform. Theory. 1978. V. 24. № 3. P. 384–386. https://doi.org/10.1109/TIT.1978.1055873
13. *Eisenbarth T., Kumar S., Paar C., Poschmann A., Uhsadel L.* A Survey of Lightweight-Cryptography Implementations // IEEE Des. Test Comput. 2007. V. 24. № 6. P. 522–533. https://doi.org/10.1109/MDT.2007.178
14. *Ivanov F., Krouk E., Kreshchuk A.* On the Lightweight McEliece Cryptosystem for Low-Power Devices // Proc. 2019 XVI Int. Symp. "Problems of Redundancy in Information and Control Systems" (REDUNDANCY). Moscow, Russia. Oct. 21–25, 2019. P. 133–138. https://doi.org/10.1109/REDUNDANCY48165.2019.9003324
15. *Misoczki R., Tillich J.-P., Sendrier N., Barreto P.S.L.M.* MDPC-McEliece: New McEliece Variants from Moderate Density Parity-Check Codes // Proc. 2013 IEEE Int. Symp. on Information Theory (ISIT'2013). Istanbul, Turkey. July 7–12, 2013. P. 2069–2073. https://doi.org/10.1109/ISIT.2013.6620590
16. *Baldi M., Chiaraluce F., Garello R., Mininni F.* Quasi-cyclic Low-Density Parity-Check Codes in the McEliece Cryptosystem // Proc. 2007 IEEE Int. Conf. on Communications (ICC'2007). Glasgow, UK. June 24–28, 2007. P. 951–956. https://doi.org/10.1109/ICC.2007.161
17. *Крук Е.А.* Граница для сложности декодирования линейных блоковых кодов // Пробл. передачи информ. 1989. Т. 25. № 3. С. 103–107. http://mi.mathnet.ru/ppi665
18. *Сидельников В.М., Шестаков С.О.* О системе шифрования, построенной на основе обобщенных кодов Рида – Соломона // Дискрет. матем. 1992. Т. 4. № 3. С. 57–63. http://mi.mathnet.ru/dm747
19. *Bernstein D.J., Lange T., Peters C.* Attacking and Defending the McEliece Cryptosystem // Post-Quantum Cryptography (Proc. 2nd Int. Workshop on Post-Quantum Cryp-

tography (PQCrypto 2008). Cincinnati, OH, USA. Oct. 17–19, 2008). Lect. Notes Comput. Sci. V. 5299. Berlin: Springer, 2008. P. 31–46. https://doi.org/10.1007/978-3-540-88403-3_3

20. *Becker A., Joux A., May A., Meurer A.* Decoding Random Binary Linear Codes in $2^{n/20}$: How $1 + 1 = 0$ Improves Information Set Decoding // Advances in Cryptology — EUROCRYPT 2012 (Proc. 31st Annu. Int. Conf. on the Theory and Applications of Cryptographic Techniques. Cambridge, UK. Apr. 15–19, 2012). Lect. Notes Comput. Sci. V. 7237. Berlin: Springer, 2012. P. 520–536. https://doi.org/10.1007/978-3-642-29011-4_31

Зяблов Виктор Васильевич
Институт проблем передачи информации
им. А.А. Харкевича РАН
zyablov@iitp.ru
Иванов Федор Ильич
Институт проблем передачи информации
им. А.А. Харкевича РАН
Национальный исследовательский университет
"Высшая школа экономики"
fivanov@hse.ru
Крук Евгений Аврамович
Национальный исследовательский университет
"Высшая школа экономики"
ekrouk@hse.ru
Сидоренко Владимир Рэмович
Институт проблем передачи информации
им. А.А. Харкевича РАН
Технический университет Мюнхена, Германия
vladimir.sidorenko@tum.de

Поступила в редакцию
30.09.2020
После доработки
14.04.2022
Принята к публикации
16.04.2022