ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 2021 Вып. 2

УДК 621.391 : 519.72

© 2021 г. **Е.Е. Егорова, Г.А. Кабатянский**

РАЗДЕЛИМЫЕ КОДЫ ДЛЯ ЗАЩИТЫ МУЛЬТИМЕДИА ОТ НЕЛЕГАЛЬНОГО КОПИРОВАНИЯ КОАЛИЦИЯМИ[1]

Дается обзор известных результатов о кодах, способных защитить мультимедийный контент от нелегального перераспределения коалициями недобросовестных пользователей.

Ключевые слова: разделимый код, разделяющий код, дизъюнктивный код, мультимедийный код цифровых отпечатков пальцев, канал множественного доступа, сигнатурный код, целенаправленный шум.

DOI: 10.31857/S0555292321020066

§ 1. Введение

Развитие Интернета и мультимедийных технологий сделало актуальной задачу разработки методов защиты цифровых авторских прав и предотвращения нелегальной перепродажи мультимедийного контента коалициями недобросовестных пользователей. Впервые математическая модель этой проблемы была сформулирована около двадцати лет назад в статье [1], а затем и в монографии [2]. Ключевым моментом предложенной модели было использование шумоподобных сигналов для построения уникальных меток, называемых цифровыми водяными знаками и внедряемых в распространяемые копии мультимедийного файла без потери качества (звука, изображения и т.д.). Вслед за этим последовал ряд работ, в которых эта модель получила свое дальнейшее развитие. В том числе, в работах [3,4] было введено понятие *разделимого* (separable) *кода*, который позволяет безошибочно находить по нелегальной копии *всех членов коалиции*. Проблему защиты мультимедийного контента от коалиционных атак можно рассматривать как обобщение на непрерывный случай другой известной задачи – о кодах цифровых отпечатков пальцев, устойчивых к коалиционным атакам (см. [5–13]). Отметим существенное различие разделимых кодов от кодов цифровых отпечатков пальцев, так как первые гарантируют нахождение всей коалиции, тогда как последние позволяют найти только одного члена коалиции (см. обзор [14]).

Построенные в [3] первые разделимые коды позволяли по нелегальной копии находить целиком коалицию недобросовестных пользователей, однако основной недостаток этих кодов был в том, что их скорость стремилась к нулю с ростом длины кода, т.е. эти коды не обеспечивали защиту мультимедийной информации с экспоненциальным от длины цифровых водяных знаков, т.е. от длины кода, числом пользователей. Этот недостаток был впервые преодолен в работе [15], где были построены разделимые коды с экспоненциальным числом пользователей, в том числе разделимые коды с простым “декодированием”, т.е. нахождением всех участников коалиции. Эти результаты были получены с помощью установленной связи между разделимы-

[1] Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (номера проектов 20-17-50013 и 20-51-50007).

ми кодами и сигнатурными кодами для специального типа канала множественного доступа, известного как A-канал [16]. В дальнейшем эта связь была расширена на другие модели защиты мультимедийной информации и соответствующие им каналы множественного доступа и стала основным теоретико-информационным методом в исследовании разделимых кодов. В частности, в работе [17] была предложена более общая математическая модель построения цифровых водяных знаков, названная авторами взвешенным двоичным суммирующим каналом множественного доступа, который обобщает как обычный двоичный суммирующий канал (см. [18]), так и его расширение, предложенное в [19]. Возникающие в этой более общей модели задачи оказались весьма близки к задачам "сжатия отсчетов" (compressed sensing), см. [20–22].

Следует отметить, что модель цифровых водяных знаков, предложенная в [1, 2], довольно чувствительна к шуму, и вопрос о том, чтобы разделимые коды были способны находить членов коалиции и в условиях шума, ставился уже в [1]. Решения для вероятностной модели были предложены в [23], а для целенаправленного шума – в [17, 24]. Отметим хорошо известную связь между кодами для каналов множественного доступа и различными вариациями комбинаторных задач поиска (см. [25]), в частности, с задачами поиска в присутствии шума, например, с игрой Реньи – Улама [26,27], известной также как задача о "поиске со лжецом". Впервые эту задачу сформулировал А. Реньи [26], но популярной она стала после книги С. Улама [27], где он задал вопрос, чему равно минимальное число вопросов, достаточное, чтобы найти неизвестное целое число в диапазоне от 1 до миллиона, если среди ответов ДА/НЕТ на вопросы один может быть ложным. Точный ответ для игры Реньи – Улама при адаптивном поиске был получен в [28]. В случае фиксированного числа L ложных ответов известна асимптотика минимального числа вопросов, а именно $(\log_2 N + L \log_2 \log_2 N)(1 + o(1))$ неадаптивных вопросов, где N – число "предметов", из которых ищется один "загаданный".

Коды по определению дискретны, а исходная постановка задачи непрерывна. Чтобы перейти в дискретную область, используется простая двоичная модуляция. А что можно улучшить, если рассмотреть более общие типы модуляции? Этот вопрос был задан и частично решен в недавней работе [29].

Наконец, в качестве лингвистического курьеза отметим, что основным инструментом для построения эффективных *разделимых* (separable) кодов стали хорошо известные в теории кодирования *разделяющие* (separating) коды, исследованию которых были посвящены многочисленные работы Ю.Л. Сагаловича и Ж. Коэна (см. их обзоры [30, 31]).

Все вышеперечисленные и некоторые другие известные результаты будут отражены в данном обзоре.

§ 2. Математические модели кодов для защиты мультимедийной информации

Математическая постановка задачи защиты цифрового контента от нелегального копирования и перераспределения возникла в конце прошлого века, см. [5–7]. Первой появилась математическая модель, наиболее известная как коды поиска пиратов [7] или коды цифровых отпечатков пальцев [11]. В этой модели имеется код над некоторым конечным алфавитом, каждому пользователю на этапе инициализации системы передается соответствующее ему кодовое слово, позволяющее получить доступ к передаваемой информации, которая зашифрована. Коалиция из не более чем t недобросовестных пользователей на основе имеющихся у ее членов кодовых слов может создать новое, ложное слово, которое тоже позволит получить доступ к зашифрованной информации. При этом имеется ограничение на то, какие слова может создавать коалиция, известное как marking assumption и существующее в двух вариациях, под названием (в терминах [12]) узкая [9] и широкая [11] выпуклые обо-

лочки. В обеих вариациях, если все члены коалиции имеют в данной позиции один и тот же символ алфавита, то он же будет стоять и в ложном слове. Если же в данной позиции не все члены коалиции одинаковы, то в случае широкой выпуклой оболочки коалиция может поставить в данной позиции ложного слова любой символ алфавита, а в случае узкой выпуклой оболочки – только один из символов, имеющихся у членов коалиции в этой позиции. Краткое изложение того, как в этой модели используется широковещательное шифрование на базе различных схем разделения секрета [32,33], можно найти в [34, Приложение]. Известные модели кодов цифровых отпечатков пальцев являются дискретными, и непрерывная модель впервые возникла в задачах защиты мультимедийного контента (изображения, музыка и т.д.) [1,2].

Рассмотрим математическую модель защиты мультимедийного контента от нелегального перераспределения. Мультимедийное сообщение представляется как N-мерный вещественный вектор $\boldsymbol{x} \in \mathbb{R}^N$. Это сообщение одновременно передается (продается) многим (M) пользователям системы. Перед передачей система уникально для каждого пользователя видоизменяет $\boldsymbol{x}$ таким образом, что если коалиция недобросовестных пользователей подделает $\boldsymbol{x}$, то система может найти всех членов коалиции. Для этого выбираются m ортонормированных векторов $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$ в $\mathbb{R}^N$, которые *не известны пользователям*. Затем система формирует для j-го пользователя свой *цифровой водяной знак* $\boldsymbol{w}_j$ как линейную комбинацию векторов $\boldsymbol{f}_i$ с двоичными коэффициентами $h_{ij} \in \{0, 1\}$:

$$\boldsymbol{w}_j = \sum_{i=1}^{m} h_{ij} \boldsymbol{f}_i. \tag{1}$$

Вложение цифровых водяных знаков осуществляется аддитивно, т.е. система выдает j-му пользователю вектор

$$\boldsymbol{y}_j = \boldsymbol{x} + \boldsymbol{w}_j \tag{2}$$

как копию $\boldsymbol{x}$, где предполагается, что длина вектора $\boldsymbol{x}$ много больше длины $\boldsymbol{w}_j$, для того чтобы копия $\boldsymbol{y}_j$ мало отличалась от оригинала $\boldsymbol{x}$.

Отметим, что известен и другой вариант “модуляции”, когда в качестве коэффициентов h_{ij} используются $+1$ и -1.

Пусть среди M пользователей системы, которым присвоены номера $1, \ldots, M$, имеется коалиция $A \subset \{1, \ldots, M\}$ недобросовестных пользователей. *Линейная атака* состоит в том, что коалиция A генерирует поддельную копию $\boldsymbol{y}$ как линейную комбинацию имеющихся у нее копий $\boldsymbol{y}_j$ с вещественными коэффициентами $\lambda_1, \ldots, \lambda_M$, такими что $\sum\limits_{j=1}^{M} \lambda_j = 1$, $\lambda_j > 0$ для всех $j \in A$ и $\lambda_j = 0$ для $j \notin A$, т.е.

$$\boldsymbol{y} = \sum_{j=1}^{M} \lambda_j \boldsymbol{y}_j = \sum_{a \in A} \lambda_a \boldsymbol{y}_a, \tag{3}$$

Так как $\sum\limits_{j=1}^{M} \lambda_j = 1$, то $\boldsymbol{y} = \boldsymbol{x} + \sum\limits_{j=1}^{M} \lambda_j \boldsymbol{w}_j$, при этом все $\lambda_j \geqslant 0$, и поэтому в силу неравенства треугольника (для нормы)

$$\|\boldsymbol{y} - \boldsymbol{x}\| = \left\| \sum_{j=1}^{M} \lambda_j \boldsymbol{w}_j \right\| \leqslant \sum_{j=1}^{M} \lambda_j \|\boldsymbol{w}_j\| \leqslant \max_j \|\boldsymbol{w}_j\| \ll \|\boldsymbol{x}\|, \tag{4}$$

где здесь и ниже $\|\boldsymbol{a}\| := \sqrt{\sum\limits_{i=1}^{N} a_i^2}$ обозначает евклидову норму вектора $\boldsymbol{a}$.

Следовательно, $\boldsymbol{y}$ является достаточно хорошей копией оригинала $\boldsymbol{x}$. Отметим, что общепринято рассматривать только линейные атаки, так как для известных примеров нелинейных атак [35] $\boldsymbol{y}$ не является достаточно хорошей копией оригинала $\boldsymbol{x}$.

Так как система знает значение $\boldsymbol{x}$, то для определения того, что $\boldsymbol{y}$ – нелегальная копия, и нахождения всех участников коалиции, которые создали $\boldsymbol{y}$, система вычисляет скалярные произведения

$$s_k = (\boldsymbol{y} - \boldsymbol{x}, \boldsymbol{f}_k) = \left(\sum_{j=1}^{M} \lambda_j \sum_{i=1}^{m} h_{ij} \boldsymbol{f}_i, \boldsymbol{f}_k \right) = \sum_{j=1}^{M} \lambda_j h_{kj} = \sum_{j \in A} \lambda_j h_{kj}, \tag{5}$$

из которых формирует вектор-*синдром*

$$\boldsymbol{S} = \boldsymbol{S}(\Lambda) = (s_1, \ldots, s_m), \tag{6}$$

где $\Lambda = (\lambda_1, \ldots, \lambda_M)$. Отметим, что носитель $\operatorname{supp}(\Lambda) := \{j : \lambda_j \neq 0\}$ вектора Λ – это и есть коалиция A.

Введем векторы $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M$, где $\boldsymbol{h}_j = (h_{1j}, \ldots, h_{mj})$. Тогда (5) можно переписать в виде

$$\boldsymbol{S}(\Lambda) = \sum_{j=1}^{M} \lambda_j \boldsymbol{h}_j = \sum_{a \in A} \lambda_a \boldsymbol{h}_a. \tag{7}$$

Это уравнение, в свою очередь, можно записать как матричное уравнение

$$\boldsymbol{S}(\Lambda) = H\Lambda^T, \tag{8}$$

где H – $(m \times M)$-матрица, составленная из векторов-столбцов $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M$.

Так как векторы $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$ ортонормированные, а векторы $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M$ выражаются в базисе $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$ как $\boldsymbol{w}_j = \sum\limits_{i=1}^{m} h_{ij} \boldsymbol{f}_i$, где $h_{ij} \in \{0, 1\}$, то множества $\mathcal{W} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M\} \subset \mathbb{R}^N$ и $\mathcal{H} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M\} \subset \{0, 1\}^m \subset \mathbb{R}^m$ изометричны. Поэтому далее мы будем оба множества называть мультимедийным кодом, а если по синдрому $\boldsymbol{S}$ можно однозначно найти носитель $\operatorname{supp}(\Lambda)$, т.е. коалицию A, то будем называть такой код *мультимедийным кодом со свойством полного поиска t-коалиций*, сокращенно – t-МППК-кодом [24] (английский эквивалент этого названия – complete traceability code [36]).

Определение 1. Двоичный код $\mathcal{H} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M\} \subset \{0, 1\}^m := B^m$ длины m называется *t-МППК-кодом*, если для любых двух вещественных векторов $\Lambda = (\lambda_1, \ldots, \lambda_M)$ и $\Lambda' = (\lambda'_1, \ldots, \lambda'_M)$, таких что все λ_j и λ'_j неотрицательны, $\sum\limits_{j=1}^{M} \lambda_j = \sum\limits_{j=1}^{M} \lambda'_j = 1$ и $|\operatorname{supp}(\Lambda)|, |\operatorname{supp}(\Lambda')| \leqslant t$, из $\operatorname{supp}(\Lambda) \neq \operatorname{supp}(\Lambda')$ следует, что $H\Lambda^T \neq H\Lambda'^T$.

Замечание 1. В данном определении условие $\operatorname{supp}(\Lambda) \neq \operatorname{supp}(\Lambda')$ можно заменить на $\operatorname{supp}(\Lambda) \cap \operatorname{supp}(\Lambda') = \varnothing$, т.е. ограничиться случаем, когда соответствующие коалиции не пересекаются.

Чтобы показать справедливость этого замечания, предположим противное, т.е. что определение выполнено для всех векторов, таких что их носители не пересекаются, но тем не менее существуют два вектора Λ и Λ', удовлетворяющие условиям определения, такие что их синдромы равны $H\Lambda^T = H\Lambda'^T$, хотя $\operatorname{supp}(\Lambda) \neq \operatorname{supp}(\Lambda')$.

Обозначим $\operatorname{supp}\Lambda = U$, $\operatorname{supp}\Lambda' = V$, $W = U \cap V$, $W_U = \{w \in W : \lambda_w > \lambda'_w\}$ и $W_V = \{w \in W : \lambda'_w > \lambda_w\}$. Условие $H\Lambda^T = H\Lambda'^T$ представимо в виде

$$\sum_{\boldsymbol{u}\in U} \lambda_u \boldsymbol{u} = \sum_{\boldsymbol{v}\in V} \lambda'_v \boldsymbol{v}. \tag{9}$$

С помощью множеств W_U и W_V перепишем (9) в виде

$$\sum_{\boldsymbol{u}\in U\setminus W} \lambda_u \boldsymbol{u} + \sum_{\boldsymbol{u}\in W_U} (\lambda_u - \lambda'_u)\boldsymbol{u} = \sum_{\boldsymbol{v}\in V\setminus W} \lambda'_v \boldsymbol{v} + \sum_{\boldsymbol{v}\in W_V} (\lambda'_v - \lambda_v)\boldsymbol{v}. \tag{10}$$

Обозначим через σ_L и σ_R суммы коэффициентов в левой и правой частях равенства (10) соответственно. Легко проверить, что $\sigma_L = \sigma_R := \sigma \leqslant 1$ и что все коэффициенты в (10) положительны. Следовательно, пронормировав коэффициенты, поделив их на σ, получим два вектора, синдромы которых совпадают, тогда как их носители, равные $(U \setminus W) \cup W_U$ и $(V \setminus W) \cup W_V$, не пересекаются, в противоречии с исходным предположением.

Пример. Проверим, что код $\mathcal{H} = \{\boldsymbol{h}_1 = (1,0), \boldsymbol{h}_2 = (0,1), \boldsymbol{h}_3 = (1,1)\}$ является 2-МППК-кодом. Заметим, что у вектора, полученного как линейная комбинация векторов $\boldsymbol{h}_1$ и $\boldsymbol{h}_2$, сумма координат равна 1, а у двух других линейных комбинаций $\lambda_i \boldsymbol{h}_i + \lambda_3 \boldsymbol{h}_3$, где $i \in \{1,2\}$, сумма координат равна $1 + \lambda_3$. Осталось проверить, возможно ли, что $\lambda_1 \boldsymbol{h}_1 + \lambda_3 \boldsymbol{h}_3 = \lambda'_2 \boldsymbol{h}_2 + \lambda'_3 \boldsymbol{h}_3$. В этом случае, как только что было отмечено, $1 + \lambda_3 = 1 + \lambda'_3$, т.е. $\lambda_3 = \lambda'_3$. Следовательно, $\lambda_1 \boldsymbol{h}_1 = \lambda'_2 \boldsymbol{h}_2$, откуда $\lambda_1 = \lambda'_2 = 0$, что противоречит предположению, что все λ положительны.

Обозначим через $M(t,m)$ максимальную мощность двоичного t-МППК-кода длины m. Из примера следует, что $M(2,2) \geqslant 3$, а так как полный код, очевидно, не является 2-МППК-кодом, то $M(2,2) = 3$.

Для оценки асимптотического поведения максимальной мощности $M(t,m)$ двоичного t-МППК-кода длины m рассмотрим, как обычно, соответствующую скорость кода $R(m,t) := m^{-1}\log_2 M(m,t)$. Будем обозначать через $R^*(t)$ и $R_*(t)$, соответственно, верхний и нижний пределы величины $R(m,t)$ при $m \to \infty$. Возьмем в определении 1 в качестве Λ и Λ' векторы, у которых координаты носителя вектора одинаковы и равны $1/t$ (такая линейная атака называется атакой усреднения, и мы ее обсудим чуть ниже). Отсюда следует, что для любого t-МППК-кода все суммы по t векторов кода (как вещественных векторов) различны и поэтому справедлив следующий аналог границы Хэмминга:

$$C^t_{M(m,t)} \leqslant (t+1)^m.$$

Тем самым, $R^*(t) \leqslant t^{-1}\log_2(t+1)$. Отметим, что при больших t эту границу можно в два раза улучшить, см. [37].

С другой стороны, в [17] была доказана конструктивно следующая нижняя граница:

$$M(m,t) \geqslant 2^{\lfloor m/t \rfloor}, \tag{11}$$

из которой очевидно следует, что $R_*(t) \geqslant t^{-1}$. Явное построение в [17] t-МППК-кодов основано на простом замечании из линейной алгебры, что если некоторые двоичные векторы $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_L$ линейно независимы над полем $GF(2)$ из двух элементов, то эти векторы линейно независимы и над полем $\mathbb{R}$ вещественных чисел. Следовательно, множество столбцов проверочной матрицы любого двоичного кода с расстоянием $d \geqslant 2t+1$ является t-МППК-кодом. Такой t-МППК-код по известному "синдрому" $S(\Lambda)$ позволяет найти вектор Λ целиком, а не только его носитель. Граница (11)

получается, если взять в качестве двоичного кода неприводимый код Гоппы длины $n = 2^\ell$ с избыточностью $r = t\ell$ и расстоянием $d \geqslant 2t + 1$, см. [38].

Замечание 2. Приведенное выше построение t-МППК-кодов не использует неотрицательности коэффициентов λ, так как в результате построения получается подмножество вершин булева куба, таких что любые $2t$ из них линейно независимы над $\mathbb{R}$. В связи с этим в [17] был задан следующий вопрос: чему равна максимально возможная мощность $M^*(t, m)$ подмножеств вершин булева куба $B^m \subset \mathbb{R}^m$, таких что любые $2t$ из них линейно независимы над $\mathbb{R}$, и какова асимптотика величины $M^*(t, m)$ при фиксированном t и $m \to \infty$? Очевидно, что $M(t, m) \geqslant M^*(t, m)$. Оказалось, что логарифмическая асимптотика $M^*(t, m)$ известна благодаря работе [39]. Обозначим через $r(n, t)$ минимальную размерность евклидова пространства, в котором существует n двоичных векторов, таких что любые t из них линейно независимы. В [39] было доказано, что $r(n, t) = O\Big(t + \dfrac{t \log(t^{-1} n)}{\log t}\Big)$, откуда, в частности, следует, что

$$c_1 \frac{\log t}{t} \leqslant R_*(t) \leqslant R^*(t) \leqslant c_2 \frac{\log t}{t}, \tag{12}$$

где $0 < c_1 < c_2$ – некоторые константы.

Среди всех линейных атак принято особо выделять *атаку усреднения*, для которой $\lambda_j = |A|^{-1}$ при $j \in A$ и $\lambda_j = 0$ в противном случае, где A – коалиция недобросовестных пользователей. Начиная с первых работ по этой тематике (см. [1, 2]), считалось, что это самая эффективная из всех линейных атак, которой можно "заменить" все остальные линейные атаки. Так, например, в [3] написано: "атака усреднения является наиболее справедливой для участников коалиции, чтобы избежать обнаружения", и поэтому в подавляющем большинстве работ ограничивались рассмотрением только атаки усреднения. Довольно очевидно, что такое утверждение неверно, так как выбор какой-то одной стратегии задания коэффициентов λ_j из всех возможных линейных атак существенно упрощает задачу поиска коалиции. Сейчас мы покажем, что в случае $t > 2$ атака усреднения не является оптимальной и среди фиксированных линейных атак.

Начнем с рассмотрения случая $t = 2$, когда, как мы сейчас покажем, *кроме атаки усреднения, нет других нетривиальных атак*. Более точно, покажем, что для любых четырех различных двоичных векторов $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}'$ из того, что

$$\lambda_a \boldsymbol{a} + \lambda_b \boldsymbol{b} = \lambda'_a \boldsymbol{a}' + \lambda'_b \boldsymbol{b}' := \boldsymbol{S} = (s_1, \ldots, s_m),$$

следует, что все λ равны $1/2$. Введем следующие множества координат: $N_{\alpha,\beta} = \{i : a_i = \alpha,\ b_i = \beta\}$ и $N'_{\alpha,\beta} = \{i : a'_i = \alpha,\ b'_i = \beta\}$, где $\alpha, \beta \in \{0, 1\}$. Очевидно, что $s_i = 0$ для $i \in N_{0,0}$ и $s_i = 1$ для $i \in N_{1,1}$, так как $\lambda_a + \lambda_b = 1$. Аналогично, $s_i = 0$ для $i \in N'_{0,0}$ и $s_i = 1$ для $i \in N'_{1,1}$. Следовательно, $N_{0,0} = N'_{0,0}$ и $N_{1,1} = N'_{1,1}$. Далее, $s_i = \lambda_a$ для $i \in N_{1,0}$ и $s_i = \lambda_b$ для $i \in N_{0,1}$. Аналогично, $S_i = \lambda'_a$ для $i \in N'_{1,0}$ и $S_i = \lambda'_b$ для $i \in N'_{0,1}$. Напомним, что $0 < \lambda_a, \lambda_b < 1$, и пусть $\lambda_a \neq \lambda_b$. Если среди значений S_i есть два значения, отличных от 0 и 1, то эти значения равны λ_a и λ_b и, аналогично, равны λ'_a и λ'_b. Следовательно, множества $\{\lambda_a, \lambda_b\}$ и $\{\lambda'_a, \lambda'_b\}$ совпадают. Пусть для простоты $\lambda_a = \lambda'_a$, $\lambda_b = \lambda'_b$, и значит, совпадают и соответствующие множества координат $N_{\alpha,\beta}$ и $N'_{\alpha,\beta}$ при всех $\alpha, \beta \in \{0, 1\}$. Тем самым, пары векторов $\boldsymbol{a}, \boldsymbol{b}$ и $\boldsymbol{a}', \boldsymbol{b}'$ совпадают. Аналогично рассматривается и случай, когда среди значений S_i имеется только одно значение, отличное от 0 и 1.

Замечание 3. Доказанное выше утверждение можно сформулировать геометрически следующим образом: любые два отрезка с вершинами в точках булева куба B^m пересекаются либо в вершинах, либо в серединах отрезков.

Приведем пример, что для больших t это уже не так. Рассмотрим $t = 3$ и двоичный код $\mathcal{H}$, состоящий из векторов $a = (1,0,0,0)$, $b = (0,1,0,1)$, $c = (0,1,1,0)$, $a' = (0,1,0,0)$, $b' = (1,0,0,1)$, $c' = (1,0,1,0)$, и две коалиции $I = \{a,b,c\}$ и $I' = \{a',b',c'\}$. Тогда при атаке усреднения система различает коалиции I и I', так как у коалиций получаются разные синдромы: $S_I = (1/3, 2/3, 1/3, 1/3)$ и $S_{I'} = (2/3, 1/3, 1/3, 1/3)$. С другой стороны, выборы $\lambda_a = 1/2$, $\lambda_b = 1/4$, $\lambda_c = 1/4$ и $\lambda_{a'} = 1/2$, $\lambda_{b'} = 1/4$, $\lambda_{c'} = 1/4$ дают один и тот же синдром $S = (1/2, 1/2, 1/4, 1/4)$. Следовательно, при такой атаке система не может различить коалиции I и I'.

Перейдем теперь к *дискретной* версии описанной выше модели. Вся полезная информация, которую система имеет о ложном векторе, содержится в скалярных произведениях s_k, см. (5). Из последнего равенства в уравнении (5) и того, что $\lambda_j > 0$ для всех $j \in A$, следует, что

$s_k = 0$, если $h_{kj} = 0$ для всех $j \in A$,

$s_k = 1$, если $h_{kj} = 1$ для всех $j \in A$,

$0 < s_k < 1$ в противном случае.

В работах [1, 3] было предложено рассматривать следующую дискретную модель, при которой системе известно не точное значение s_k, а только то, что $s_k = 0$, $s_k = 1$, или $0 < s_k < 1$. Это равносильно тому, что системе известно, что либо все k-е координаты векторов коалиции равны 0, что соответствует случаю $s_k = 0$, либо все k-е координаты векторов коалиции равны 1 (случай $s_k = 1$), либо среди значений k-й координаты встречаются как 0, так и 1.

Обозначим через $F(\cdot)$ следующее отображение отрезка $[0,1]$ на троичный алфавит, состоящий из символов $0, 1$ и $*$:

$$F(x) = \begin{cases} x, & \text{если } x \in \{0,1\}, \\ *, & \text{если } 0 < x < 1, \end{cases}$$

где $*$ используется для краткости как символ, заменяющий множество $\{0,1\}$.

Для произвольного множества A вершин булева куба B^m определим его проекцию на i-ю координату как $P_i(A) := \{a_i : \boldsymbol{a} \in A\}$ и его полную проекцию как

$$P(A) = \{(a_1, \ldots, a_m) : \ a_1 \in P_1(A), \ldots, a_m \in P_m(A)\}, \tag{13}$$

т.е. $P(A) = P_1(A) \times \ldots \times P_m(A)$.

Для произвольного множества (коалиции) A его *дискретный синдром* $F(S) = (F(s_1), \ldots, F(s_m)) = F(A)$ *не зависит от выбора коэффициентов* λ_a *в* (5), так как $F(s_i) = P_i(A)$ в силу того, что $\lambda_a > 0$ при $a \in A$ и $\lambda_a = 0$ в противном случае. Тем самым, задачей системы становится восстановить коалицию по ее дискретному синдрому или, что то же самое, по ее полной проекции. Это приводит к определению разделимого (separable) кода.

О п р е д е л е н и е 2 [3, 4]. Двоичный код C называется *t-разделимым кодом*, если для любых двух различных кодовых подмножеств (коалиций) $U, V \subset C$, $|U| \leqslant t$, $|V| \leqslant t$, их полные проекции различны:

$$P(U) \neq P(V). \tag{14}$$

Сравним определения 1 и 2. Так как для любой коалиции ее дискретный синдром не зависит от коэффициентов λ, то t-разделимый код является t-МППК-кодом. Обратное в общем случае неверно, так как нижняя оценка (12) для скорости t-МППК-кодов превышает верхнюю оценку (23) на скорость t-разделимых кодов для больши́х t. Отметим также, что в определении 2 нельзя заменить условие "различные подмножества" на "непересекающиеся", как можно было сделать в определении 1.

Однако для $t = 2$ эти два понятия совпадают. Для этого покажем, что 2-МППК-код является 2-разделимым. Рассмотрим четыре произвольных кодовых вектора $\boldsymbol{a} \neq \boldsymbol{b} \neq \boldsymbol{c} \neq \boldsymbol{d}$ и соответствующие две коалиции $U = \{\boldsymbol{a}, \boldsymbol{b}\}$ и $V = \{\boldsymbol{c}, \boldsymbol{d}\}$ и применим к ним атаку усреднения. Так как код 2-МППК, то $\boldsymbol{a}+\boldsymbol{b} \neq \boldsymbol{c}+\boldsymbol{d}$, что в силу двоичности векторов равносильно тому, что $P(V) \neq P(U)$, что и требовалось доказать.

Определение 2 показывает и сходство, и различие между разделимыми кодами и идентифицирующими кодами [14], также известными как коды со свойством отождествления родителей, или IPP-коды [9]. Напомним, что q-ичный код C называется *t-IPP-кодом*, если для любого вектора $\boldsymbol{z}$ либо

$$\bigcap_{U\colon \boldsymbol{z}\in P(U),\, U\subset C,\, |U|\leqslant t} U \neq \varnothing, \tag{15}$$

либо не существует кодового подмножества U (коалиции), такого что $\boldsymbol{z} \in P(U)$ и $|U| \leqslant t$. Таким образом, в случае IPP-кодов по любой точке $\boldsymbol{z}$ из полной проекции $P(U)$ коалиции U система может гарантированно найти хотя бы одного члена коалиции. Отметим, что в такой постановке задачи *найти коалицию целиком невозможно*, за исключением случая тривиальных кодов, мощность которых не более мощности алфавита. Действительно, рассмотрим произвольный q-ичный код C мощности $|C| > q$. Пусть i – некоторая координата, такая что $|P_i(C)| > 1$. Так как $|C| > q$, то существуют векторы $\boldsymbol{b}, \boldsymbol{b}' \in C$, такие что $b_i = b'_i$. Так как $|P_i(C)| > 1$, то существует вектор $\boldsymbol{a} \in C$, такой что $a_i \neq b_i$. Тогда вектор $\boldsymbol{z}$, который совпадает с вектором $\boldsymbol{a}$ во всех координатах, кроме i, где $z_i = b_i$, порождается двумя разными коалициями: $\{\boldsymbol{a}, \boldsymbol{b}\}$ и $\{\boldsymbol{a}, \boldsymbol{b}'\}$.

Очень важным является следующий факт, впервые отмеченный в [15], что понятие разделимых кодов совпадает с известным в теории информации понятием сигнатурных кодов для А-канала множественного доступа. Напомним соответствующие определения.

Мы будем рассматривать детерминированные каналы множественного доступа без памяти (сокращенно, MAC – multiple access channel) с дискретным временем и частичной активностью (см. [18]). MAC задается входным и выходным алфавитами X и Y и функцией выхода $f\colon X^M \to Y$, такой что выход MAC равен $y = f(x_1, \ldots, x_M)$, где $x_j \in X \cup \varnothing$ – символ, подаваемый на вход канала j-м пользователем, причем $x_j = \varnothing$ означает, что j-й пользователь ничего не подал на вход MAC, т.е. был не активен.

Нас будут особо интересовать *t-сигнатурные* коды, когда каждому пользователю сопоставлено только одно кодовое слово, а именно j-му пользователю сопоставлено слово $\boldsymbol{c}_j$, и не более чем t пользователей активны, т.е. передают в канал свои кодовые слова. Если при этом по выходу канала можно однозначно восстановить, какие пользователи были активны, при условии, что их было не более t, то код $C = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\}$ называется *t-сигнатурным*. Далее мы будем отождествлять пользователя и соответствующее ему кодовое слово. Для активного множества пользователей U будем обозначать через S_U соответствующий выход MAC.

О п р е д е л е н и е 3. Код C называется *t-сигнатурным*, если для любых двух различных подмножеств $U, V \subset C$, таких что $|U|, |V| \leqslant t$, справедливо $S_U \neq S_V$.

Наиболее важными для нас примерами MAC являются А-канал, $\vee$-канал и двоичный суммирующий канал, см. [18].

Для булева суммирующего канала, сокращенно $\vee$-канала, $X = Y = \{0, 1\}$ и $f(x_1, \ldots, x_t) = x_1 \vee \ldots \vee x_t$.

Для двоичного суммирующего канала, сокращенно Σ-канала, $X = \{0, 1\}$, $Y = \{0, 1, 2, \ldots\}$ и

$$f(x_1, \ldots, x_t) = x_1 + \ldots + x_t \in \mathbb{Z}.$$

Для А-канала вход X – произвольное конечное множество, $Y = 2^X$ – множество всех подмножеств X и $f(x_1, \ldots, x_t) = \{x_1, \ldots, x_t\}$. Например, для $X = \{0, 1, 2\}$ имеем $f(0, 1, 1, 0) = \{0, 1\}$. Тем самым, в А-канале для любого множества $U \subset C$ справедливо $S_U = P(U)$, и следовательно, t-разделимый код – это то же самое, что t-сигнатурный код для А-канала. Этот факт впервые был отмечен в [15], и начиная с этой работы, сигнатурные коды стали интенсивно использоваться при изучении кодов для защиты мультимедийной информации.

Модель А-канала была введена в работе [16] вместе с В-каналом, выходом которого является не только множество символов, переданных активными пользователями, но и кратности, с которыми эти символы были использованы.

Введем *частичный порядок на каналах множественного доступа*. Пусть каналы $\mathcal{A}$ и $\mathcal{B}$ задаются одним и тем же входным алфавитом X, множествами выходов $Y_{\mathcal{A}}$ и $Y_{\mathcal{B}}$ и функциями выхода $f_{\mathcal{A}}$ и $f_{\mathcal{B}}$. Будем говорить, что канал $\mathcal{A}$ меньше, чем канал $\mathcal{B}$, и обозначать $\mathcal{A} \prec \mathcal{B}$, если существует отображение $g\colon Y_{\mathcal{B}} \to Y_{\mathcal{A}}$, такое что для любого набора $x_1, \ldots, x_m$ справедливо

$$g(f_{\mathcal{B}}(x_1, \ldots, x_m)) = f_{\mathcal{A}}(x_1, \ldots, x_m). \tag{16}$$

Соотношение (16) говорит, что "больший" канал $\mathcal{B}$ дает "больше информации", чем канал $\mathcal{A}$, а именно что по выходу канала $\mathcal{B}$ можно однозначно восстановить выход канала $\mathcal{A}$. Отсюда следует, что если $\mathcal{A} \prec \mathcal{B}$ и код $C \subset X^n$ является t-сигнатурным кодом для канала $\mathcal{A}$, то он является t-сигнатурным кодом и для канала $\mathcal{B}$.

В двоичном случае, т.е. при $X = \{0, 1\}$, имеет место следующее упорядочение: $\vee$-канал $\prec$ A-канал $\prec$ B-канал.

Отметим, что в работе [3] было показано, что t-дизъюнктивные (t-superimposed) коды, введенные в [40], являются t-разделимыми кодами. Действительно, t-дизъюнктивный код является t-сигнатурным кодом для $\vee$-канала, а так как $\vee$-канал "меньше" А-канала, то t-дизъюнктивный код является t-сигнатурным кодом для А-канала, т.е. является t-разделимым кодом. Поэтому условие дизъюнктивности является достаточным для разделимости, но не необходимым. Мы обсудим это подробнее в следующем параграфе.

§ 3. Разделимые, разделяющие и дизъюнктивные коды

Как было объяснено в предыдущем параграфе, разделимые коды – это то же самое, что сигнатурные коды для *двоичного* А-канала множественного доступа. Так как А-канал определен для любого конечного алфавита, то естественно обобщить определение 3 разделимого кода на случай произвольного алфавита.

Пусть X – конечный алфавит мощности q. Для произвольного множества $A \subset X^m$ обозначим через

$$P(A) = \{(x_1, \ldots, x_m) \in X^m : \; x_1 \in P_1(A), \; \ldots, \; x_m \in P_m(A)\} \tag{17}$$

его полную проекцию, где $P_i(A) := \{a_i : \boldsymbol{a} \in A\}$ – это проекция множества A на i-ю координату. Таким образом, $P(A) = P_1(A) \times \ldots \times P_m(A)$.

О п р е д е л е н и е 4. Код $C \subset X^m$ называется *t-разделимым*, если для любых двух различных подмножеств $U, V \subset C$, таких что $|U|, |V| \leqslant t$, справедливо $P(U) \neq P(V)$.

Таким образом, для разделимого кода любое его подмножество мощности не более t может быть однозначно найдено по своей полной проекции. Как уже отмечалось, свойство t-разделимости кода совпадает со свойством кода быть t-сигнатурным для А-канала, т.е. позволяет по выходу А-канала однозначно восстановить, какие пользователи были активны.

Исторически еще раньше, в 60-е годы прошлого века [41], появилось понятие *разделяющего* кода, см. [30].

Определение 5. q-ичный код C называется (t_1, t_2)-разделяющим, если для любых двух непересекающихся подмножеств $U, V \subset C$, таких что $|U| \leqslant t_1$, $|V| \leqslant t_2$, существует координата i, которая их разделяет, т.е. $P_i(U) \cap P_i(V) = \varnothing$.

Следующее утверждение, доказанное в [3], показывает, что понятия t-разделимого кода и $(1, t)$-разделяющего кода близки.

Предложение 1. *$(1, t)$-разделяющий код является t-разделимым, а t-разделимый код является $(1, t-1)$-разделяющим.*

Доказательство. Очевидно, что для любого $(1, t)$-разделяющего кода C его произвольное подмножество (коалиция) $U \subset C : |U| \leqslant t$ может быть найдено по своей полной проекции $P(U)$ следующим образом:

$$U = \{\boldsymbol{c} \in C : \boldsymbol{c} \in P(U)\}. \tag{18}$$

Покажем теперь, что t-разделимый код C является $(1, t-1)$-разделяющим. Рассмотрим произвольное кодовое подмножество U мощности не более $t-1$ и произвольное кодовое слово $\boldsymbol{a} \notin U$ и сформируем кодовое множество $V = U \cup \boldsymbol{a}$. Тогда в силу t-разделимости кода должно быть $P(U) \neq P(V)$, и следовательно, существует координата i, такая что $a_i \notin P_i(U)$, что и требовалось доказать. ▲

Замечание 4. Частные случаи разделяющих кодов были переоткрыты в работах по кодам цифровых отпечатков пальцев: $(1, t)$-разделяющие коды под именем t-frameproof codes, а (t, t)-разделяющие коды – под именем t-secure frameproof codes [11]. Новая терминология не принесла новых результатов, за исключением исследования нетрадиционного для теории кодирования случая, когда мощность алфавита растет, а длина кода фиксирована (см., например, [42]).

Обозначим через $A^q_{\text{sep}}(t, n)$ максимально возможную мощность q-ичного t-разделимого кода длины n, а через $A^q_s(t, n)$ – максимально возможную мощность q-ичного $(1, t)$-разделяющего кода длины n. Тогда из предложения 1 следует, что

$$A^q_s(t, n) \leqslant A^q_{\text{sep}}(t, n) \leqslant A^q_s(t-1, n).$$

Как обычно, мы будем опускать символ q, когда речь идет о двоичных кодах. Также будем рассматривать асимптотическое поведение мощности наилучших кодов в виде их скорости, определяемой как

$$R_{\text{sep}}(t) = \lim_{n\to\infty} n^{-1} \log_2 A_{\text{sep}}(t, n)$$

и

$$R_s(t) = \lim_{n\to\infty} n^{-1} \log_2 A_s(t, n)$$

соответственно. Мы здесь допускаем некую вольность записи, так как мы не доказываем существования соответствующих пределов (хотя вероятно, что это можно сделать аналогично тому, как это было сделано для обычных кодов в [43]). Поэтому нижеследующие границы надо рассматривать как оценки на нижний и верхний пределы соответственно.

При малых t известные результаты относительно этих двух классов кодов заметно различаются. Например, для скорости $R_s(t)$ лучших двоичных $(1, 2)$-разделяющих кодов известно, что

$$0{,}207565 \leqslant R_s(2) \leqslant 1/2, \tag{19}$$

где нижняя граница получена сравнительно недавно в [44] с помощью алгеброгеометрических кодов, что позволило улучить известный задолго до этого аналог гра-

ницы Варшамова – Гилберта (или случайного кодирования)

$$R_s(2) \geqslant 1 - 2^{-1}\log_2 3 = 0{,}207518.$$

Покажем, что при числе активных пользователей не более двух асимптотика скорости наилучших сигнатурных кодов для двоичных суммирующего канала и А-канала асимптотически совпадают. Легко проверить, что если априори известно число активных пользователей и оно не больше чем 2, то двоичные А-канал и суммирующий канал эквивалентны (т.е. совпадают при соответствующей перенумерации выходов каналов). Пусть C – это 2-сигнатурный код для суммирующего канала. Так как все слова кода C различны, то среди координат выхода А-канала для пары слов обязательно есть символ $*$, которого нет в случае одного активного пользователя, следовательно, C различает случай одного или двух активных пользователей и поэтому является 2-сигнатурным кодом и для А-канала. Пусть C – это 2-сигнатурный код для А-канала. Удлиним C на одну координату, сделав ее равной 1 для всех кодовых слов. Тогда новый код различает случай одного или двух активных пользователей в суммирующем канале и поэтому является 2-сигнатурным кодом и для суммирующего канала.

Поэтому известные границы для скорости 2-сигнатурных кодов в суммирующем канале справедливы и для двоичных 2-разделимых кодов:

$$1/2 \leqslant R_{\mathrm{sep}}(2) \leqslant 0{,}5753, \tag{20}$$

где нижняя граница – это стандартная граница случайного кодирования, а верхняя граница получена в [45].

Имеется хорошо известная связь между двоичными $(1,t)$-разделяющими кодами и t-дизъюнктивными [40] кодами.

Определение 6. Двоичный код C называется *t-дизъюнктивным*, если для любого кодового подмножества $U \subset C$ мощности не более t и произвольного кодового слова $\boldsymbol{a} \notin U$ существует координата i, такая что $a_i = 1$, тогда как $u_i = 0$ для всех $\boldsymbol{u} \in U$ (т.е. $P_i(U) = 0$).

Дизъюнктивные коды были переоткрыты в экстремальной комбинаторике под названием семейства множеств без t-покрытий [46, 47], т.е. таких семейств, что ни одно множество семейства не покрывается объединением t других множеств этого семейства.

Очевидно, что t-дизъюнктивный код C является t-сигнатурным кодом для $\vee$-канала, т.е. для любых двух различных подмножеств $U, V \subset C$, таких что $|U|, |V| \leqslant t$, справедливо

$$\bigvee_{\boldsymbol{u}\in U} \boldsymbol{u} \neq \bigvee_{\boldsymbol{v}\in V} \boldsymbol{v}. \tag{21}$$

Это следует, например, из того, что если выход $\vee$-канала при использовании t-дизъюнктивного кода C равен $\boldsymbol{S}(U) = \boldsymbol{S} = (s_1, \ldots, s_m)$, то на вход были поданы векторы $\boldsymbol{c} \in C$, такие что $u_i \leqslant s_i$ для всех i, ср. (18).

Тем самым, t-дизъюнктивные коды играют для $\vee$-канала ту же роль, что $(1,t)$-разделяющие коды для А-канала. Очевидно, что t-дизъюнктивный код является двоичным $(1,t)$-разделяющим кодом. С другой стороны, пусть C – это двоичный $(1,t)$-разделяющий код длины n. Легко видеть, что код C^* длины $2n$, состоящий из слов вида $(\boldsymbol{x}, \overline{\boldsymbol{x}})$, где $\overline{\boldsymbol{x}}$ – двоичный вектор, полученный из вектора $\boldsymbol{x}$ инвертированием всех координат, является t-дизъюнктивным кодом. Поэтому если обозначить через $R_\vee(t)$ максимальную скорость t-дизъюнктивных кодов, то выполнено следу-

ющее соотношение:

$$R_\vee(t) \leqslant R_s(t) \leqslant 2R_\vee(t). \tag{22}$$

Для больших t известны следующие асимптотические границы, которые мы приведем для двоичного случая, а общий случай подробно исследован в [48]. Итак, при больших t

$$\Theta\Big(\frac{1}{t^2}\Big) \leqslant R_s(t) \leqslant R_{\mathrm{sep}}(t) \leqslant O\Big(\frac{\log t}{t^2}\Big). \tag{23}$$

Нижняя граница в (23) была получена в [30] стандартным методом случайного кодирования с выбрасыванием (см. [49]), тогда как верхняя граница была сначала получена для дизъюнктивных кодов в [46,47,50] и только затем перенесена на $(1,t)$-разделяющие коды с помощью (22).

§ 4. Разделимые и разделяющие коды с простым декодированием и исправлением ошибок

Так как нижняя граница в (23) – это граница существования, то возникает традиционный вопрос о кодах с "простыми" (т.е. со сложностью, полиномиальной от длины кода) алгоритмами построения кода, его кодирования и декодирования. Такой класс кодов, основанный на каскадной конструкции и "мягком" декодировании каскадных кодов [51, 52], был предложен в [15].

Дадим описание каскадной конструкции, следуя [53]. Имеются два кода: q-ичный код C длины m и мощности Q, называемый внутренним кодом, и Q-ичный код W длины N и мощности M, называемый внешним кодом, и взаимно-однозначное отображение $\varphi\colon GF(Q) \to C$. Каскадный код V состоит из слов вида $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N) = \Phi(\boldsymbol{w})$, где $\boldsymbol{v}_i = \varphi(w_i) \in C$ и $\boldsymbol{w} = (w_1, \ldots, w_N) \in W$. Этот код q-ичный, длины Nm и мощности M. В каскадной конструкции свойства внутреннего и внешнего кодов часто "наследуются". Так, нам понадобится хорошо известный (и просто проверяемый) факт (см. [30]), что *каскадный код, у которого и внутренний и внешний коды являются* $(1,t)$*-разделяющими, также является* $(1,t)$*-разделяющим.*

Также будет полезно следующее простое замечание.

Предложение 2. *Если расстояние кода больше чем* $n(1-t^{-1})$*, где* n *– его длина, то код является* $(1,t)$*-разделяющим.*

Действительно, пусть это не так и существует кодовое слово $\boldsymbol{c}$ и подмножество кода $U : |U| \leqslant t$, которые не разделяются. Тогда число совпадений координат между $\boldsymbol{c}$ и словами из U не менее n, а с другой стороны, из неравенства на расстояние кода следует, что любые два кодовых слова совпадают менее чем в nt^{-1} позициях, а так как слов в U не более чем t, то число совпадений координат между $\boldsymbol{c}$ и словами из U меньше n. Это рассуждение применялось многократно, видимо, начиная с [54], в том числе для доказательства существования IPP-кодов [7].

Опишем конструкцию из [15]. В качестве внутреннего кода берется двоичный $(1,t)$-разделяющий код C длины m и мощности Q, а в качестве внешнего – код Рида – Соломона W над полем $GF(Q)$ длины $N = Q$ и размерности $K = N/t$, т.е. код со скоростью $R = 1/t$. Каскадный код V состоит из слов вида $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N) = \Phi(\boldsymbol{w})$, где $\boldsymbol{v}_i = \varphi(w_i) \in C$ и $\boldsymbol{w} = (w_1, \ldots, w_N) \in W$. Из сказанного выше следует, что этот каскадный код является двоичным $(1,t)$-разделяющим кодом длины Nm и мощности $Q^{N/t}$. Для его декодирования в A-канале мы будем использовать алгоритм Гуруcвами – Судана "мягкого" списочного декодирования кодов Рида – Соломона, предложенный в [51, 52].

Пусть символам α конечного поля $GF(Q)$ приписаны неотрицательные веса ("надежности") $r_i(\alpha)$ в зависимости от координаты $i \in \{1, \ldots, N\}$. Определим вес ("надежность") кодового вектора $\boldsymbol{w}$ как

$$r(\boldsymbol{w}) = \sum_{i=1}^{N} r_i(w_i).$$

Алгоритм [51, 52] позволяет найти за полиномиальное от N время список всех слов кода Рида – Соломона, таких что

$$r(\boldsymbol{w}) \geqslant r_{\text{crit}} = \sqrt{NR\sigma}, \tag{24}$$

где $\sigma = \sum\limits_{i=1}^{N} \sum\limits_{\alpha \in GF(Q)} r_i^2(\alpha)$, а $R = K/N$ – скорость кода Рида – Соломона.

Опишем работу алгоритма декодирования в A-канале. Пусть на вход A-канала поступили кодовые векторы $\boldsymbol{v}^1 = \Phi(\boldsymbol{w}^1)$, ..., $\boldsymbol{v}^{t'} = \Phi(\boldsymbol{w}^{t'})$, где $t' \leqslant t$. Пусть

$$\mathbf{s} = (s_{11}, \ldots, s_{1m}, s_{21}, \ldots, s_{2m}, \ldots, s_{N1}, \ldots, s_{Nm}) = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N)$$

– соответствующий выход A-канала. Первый шаг алгоритма состоит в декодировании векторов $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$ внутренним $(1, t)$-разделяющим кодом C длины m, например, полным перебором всех слов кода C, что потребует $O(Qm)$ операций. В результате будут получены некоторые подмножества $A_1, \ldots, A_N$ слов кода C, $|A_i| := t_i \leqslant t$ для всех i, и соответствующие им подмножества $H_i = \varphi^{-1}(A_i)$ символов поля $GF(Q)$.

Зададим веса следующим образом: $r_i(\alpha) = 1$, если $\alpha \in H_i$, и $r_i(\alpha) = 0$ в противном случае. Из свойства $(1, t)$-разделимости кода C следует, что i-е координаты векторов $\boldsymbol{w}^1, \ldots, \boldsymbol{w}^{t'}$ принадлежат множеству H_i при всех i, и следовательно, $r(\boldsymbol{w}^j) = N$ для всех $j = 1, \ldots, t'$. С другой стороны, так как данный код Рида – Соломона является $(1, t)$-разделяющим (в силу того, что его расстояние $d = N - K + 1 > N(1 - t^{-1})$), то для любого другого кодового слова $\boldsymbol{w}$ существует как минимум одна координата, не принадлежащая соответствующему множеству H_i, и следовательно, $r(\boldsymbol{w}) \leqslant N - 1$.

В силу (24) алгоритм Гурусвами – Судана выдаст все слова $\boldsymbol{w}^1, \ldots, \boldsymbol{w}^{t'}$ в составе списка, но, возможно, и некоторые "лишние", которые будут затем отброшены как не прошедшие проверку $r(\boldsymbol{w}) = N$. Полиномиальность построения кода, его кодирования и декодирования следует, как обычно, из того, что мощность внутреннего кода Q растет экспоненциально от длины m, следовательно, длина итогового (каскадного) кода, равная Qm, растет так же, и основной вклад в сложность декодирования составляет сложность декодирования кода Рида – Соломона, а она полиномиальна от Q.

Замечание 5. Для дизъюнктивных кодов известна другая конструкция кодов с полиномиальной сложностью построения, кодирования и декодирования кода [55]. Достоинством предложенной здесь каскадной конструкции является то, что она легко обобщается на случай, когда синдром может быть ошибочным, см. ниже.

Как мы уже отмечали, переход от непрерывной модели мультимедийных кодов, находящих коалицию недобросовестных пользователей целиком, т.е. t-МППК-кодов (см. определение 1), к дискретной модели мультимедийных кодов, также находящих коалицию целиком, т.е. к разделимым кодам (см. определения 2 и 4), неминуемо ведет к ошибкам дискретизации в силу неточности измерений. Другой возможный источник ошибок – это недобросовестные пользователи (участники коалиции). Среди возможных моделей ошибок в этом параграфе мы уделим основное внимание комбинаторной модели ошибок, когда любой из символов на выходе A-канала может быть изменен, но число изменений-ошибок заранее ограничено сверху некоторой величи-

ной T. Этот класс ошибок также часто называют целенаправленными (adversarial), имея в виду ситуацию, когда некто пытается обмануть систему и вносит произвольные ошибки, но число вносимых ошибок заранее ограничено сверху.

Ясно, что комбинаторная модель ошибок применима к любому каналу множественного доступа. Введем следующее

Определение 7. Код C называется *t-сигнатурным кодом, исправляющим T ошибок в канале множественного доступа*, если по выходу канала, искаженному не более чем в T координатах, однозначно восстанавливается множество активных пользователей мощности не более t.

Эквивалентное этому определению условие таково:

Для любых двух различных кодовых подмножеств U и V мощности не более t каждое справедливо

$$d_H(S_U, S_V) > 2T, \tag{25}$$

где d_H – расстояние Хэмминга.

Дадим соответствующее определение для разделяющих кодов, обобщающее определение 5.

Определение 8. q-ичный код C называется *(t_1, t_2)-разделяющим с исправлением T ошибок*, если для любых двух непересекающихся подмножеств $U, V \subset C$, таких что $|U| \leqslant t_1$, $|V| \leqslant t_2$, число разделяющих их координат больше $2T$, т.е.

$$|\{i:\ P_i(U) \cap P_i(V) = \varnothing\}| > 2T. \tag{26}$$

В частности, код C является *$(1, t)$-разделяющим кодом с исправлением T ошибок*, если для любого кодового вектора $\boldsymbol{c} \in C$ и любого t-множества кода U: $\boldsymbol{c} \notin U$ имеется больше чем $2T$ разделяющих их координат. Очевидно, что такой код является t-сигнатурным кодом, исправляющим T ошибок в A-канале.

Отметим, что определение 8 появилось в работе [56] из других соображений (см. также [57]). А именно было показано, как из (t_1, t_2)-разделяющего кода получить $(t_1 - 1, t_2 - 1)$-разделяющий код с довольно большим расстоянием Хэмминга, что привело к новым верхним границам на мощность (t_1, t_2)-разделяющих кодов. Эта техника получения верхних границ для мощности (t_1, t_2)-разделяющих кодов была независимо переоткрыта в [58]. Заметим, что, к сожалению, эта техника не применима к $(1, t)$-разделяющим кодам.

Обобщим предложение 2 на случай исправления ошибок.

Предложение 3. *Если расстояние кода больше чем $n(1 - t^{-1}(1 - 2\tau))$, где n – его длина, то код является $(1, t)$-разделяющим кодом с исправлением $T = \tau n$ ошибок.*

Доказательство. Рассмотрим произвольное подмножество кода $U : |U| \leqslant t$ и любое кодовое слово $\boldsymbol{c}$ не из U. Тогда число совпадений координат между $\boldsymbol{c}$ и любым словом из U меньше чем $n\dfrac{1 - 2\tau}{t}$, следовательно, общее число совпадений координат между $\boldsymbol{c}$ и U меньше чем $n(1 - 2\tau)$. Тем самым, число координат, разделяющих $\boldsymbol{c}$ и U, больше $2n\tau$. ▲

Обобщим описанные выше конструкцию и алгоритм декодирования из [15] так, чтобы исправлять $T = \tau N$ ошибок. Внутренний код оставим без изменений, т.е. возьмем двоичный $(1, t)$-разделяющий код C длины m и мощности Q, а в качестве внешнего возьмем код Рида – Соломона W над полем $GF(Q)$ длины $N = Q$ и скорости $R = \dfrac{1 - 2\tau}{t}$.

Как и в исходном алгоритме, применимом к А-каналу без ошибок, первый шаг состоит в декодировании векторов $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$ внутренним $(1, t)$-разделяющим кодом C.

Результатом декодирования являются подмножества $A_1, \dots, A_N$ слов кода C и соответствующие им подмножества $H_i = \varphi^{-1}(A_i)$ символов поля $GF(Q)$. Веса для декодирования внешнего кода оставим прежними, т.е. $r_i(\alpha) = 1$, если $\alpha \in H_i$, и $r_i(\alpha) = 0$ в противном случае.

Из свойства $(1,t)$-разделимости кода C следует, что i-я координата любого из векторов $\boldsymbol{w}^1, \dots, \boldsymbol{w}^{t'}$ принадлежит множеству H_i, если в векторе $\boldsymbol{s}_i$ не было ошибок в A-канале. Следовательно, для любого $\boldsymbol{w}^j$ справедливо $r(\boldsymbol{w}^j) \geqslant N - T = T(1-\tau)$, так как без ошибок $r(\boldsymbol{w}^j) = N$. С другой стороны, в силу предложения 3 для любого кодового слова $\boldsymbol{w}$ не из коалиции число координат, разделяющих это слово и коалицию, больше чем $2T$, а так как ошибки могли уменьшить это число максимум на T, то $r(\boldsymbol{w}) < N - T = N(1-\tau)$. При этом алгоритм декодирования Гурусвами – Судана выдаст все слова кода, для которых $r(\boldsymbol{w}) \leqslant r_{\mathrm{crit}}$, где

$$r_{\mathrm{crit}} = \sqrt{NR\sigma} \leqslant \sqrt{N\frac{1-2\tau}{t}Nt} = N\sqrt{1-2\tau} < N(1-\tau). \tag{27}$$

Следовательно, алгоритм выдаст все слова коалиции, а возможные лишние слова будут отсеяны неравенством $r(\boldsymbol{w}) \geqslant N - T = N(1-\tau)$.

§ 5. МППК-коды с исправлением ошибок

Мы уже отмечали, что при переходе от непрерывной модели t-МППК-кодов к дискретной модели разделимых кодов неточность измерений ведет к ошибкам в соответствующем дискретном канале MAC (A-канале). Сейчас нам будет важно, что неточность измерений вносит ошибки и для t-МППК-кодов. Другой возможный источник ошибок для t-МППК-кодов – это недобросовестные пользователи (участники коалиции). Мы ограничимся рассмотрением t-МППК-кодов, способных исправлять ошибки, у которых ограничена сверху норма вектора ошибки в евклидовой метрике [24]. В [17] была рассмотрена другая модель ошибок – *разреженные ошибки*, т.е. когда ошибки могут изменить не более заранее заданного числа координат в векторе-синдроме $\boldsymbol{S}(\Lambda)$ (см. уравнения (5), (7)), но зато нет никаких ограничений на величину ошибки в изменяемых координатах. Конструкция соответствующих кодов, предложенная в [17] (см. также [59,60]), по существу излагается ниже как часть построения t-ДЕД-кодов.

Будем, следуя [36], рассматривать атаку коалиции A, которая не только создает ложную копию $\boldsymbol{y} = \boldsymbol{x} + \sum\limits_{j\in A} \lambda_j \boldsymbol{w}_j \in \mathbb{R}^N$ в соответствии с моделью линейной атаки, но еще целенаправленно добавляет вектор шума $\boldsymbol{e} \in \mathbb{R}^N$, такой что $\|\boldsymbol{e}\| \leqslant \delta$, где $\|\cdot\|$ – евклидова норма на $\mathbb{R}^N$. В результате коалиция A перераспределяет копию

$$\widehat{\boldsymbol{y}} = \boldsymbol{x} + \sum_{j\in A} \lambda_j \boldsymbol{w}_j + \boldsymbol{e}, \tag{28}$$

где $\boldsymbol{w}_j = \sum\limits_{i=1}^{m} h_{ij}\boldsymbol{f}_i$ (см. (1)). Координаты синдрома $\boldsymbol{S} = \boldsymbol{S}(\Lambda) = (s_1, \dots, s_m)$ равны

$$s_k = \left(\boldsymbol{e} + \sum_{j=1}^{M} \lambda_j \sum_{i=1}^{m} h_{ij}\boldsymbol{f}_i, \boldsymbol{f}_k\right) = (\boldsymbol{e}, \boldsymbol{f}_k) + \sum_{j=1}^{M} \lambda_j h_{kj} = \varepsilon_k + \sum_{j\in A} \lambda_j h_{kj} \tag{29}$$

(см. (5)), где $\varepsilon_k = (\boldsymbol{e}, \boldsymbol{f}_k)$, и длина вектора $\varepsilon = (\varepsilon_1, \dots, \varepsilon_m)$, равная $\|\varepsilon\| = \sqrt{\sum\limits_{i=1}^{m} \varepsilon_i^2}$, не превышает длины вектора $\boldsymbol{e}$ в силу неравенства Бесселя.

Как отмечалось в начале § 2, множество $\mathcal{W} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_M\} \subset \mathbb{R}^N$ и двоичный код $\mathcal{H} = \{\boldsymbol{h}_1, \dots, \boldsymbol{h}_M\} \subset B^m$ изометричны, и далее мы будем рассматривать код $\mathcal{H}$. Двоичный код $\mathcal{H}$ естественно называть (t,δ)-*МППК-кодом со свойством полного поиска*

t-коалиций и устойчивым к δ-шуму, если по любой ложной копии $\widehat{\boldsymbol{y}} = \sum\limits_{j\in A} \lambda_j \boldsymbol{y}_j + \boldsymbol{e}$ можно однозначно найти коалицию A.

Это условие равносильно тому, что для любых двух вещественных векторов $\Lambda = (\lambda_1, \ldots, \lambda_M)$ и $\Lambda' = (\lambda'_1, \ldots, \lambda'_M)$, таких что все λ_j и λ'_j неотрицательны, $\sum\limits_{j=1}^{M} \lambda_j = \sum\limits_{j=1}^{M} \lambda'_j = 1$ и $|\operatorname{supp}(\Lambda)|, |\operatorname{supp}(\Lambda')| \leqslant t$, из $A = \operatorname{supp}(\Lambda) \neq B = \operatorname{supp}(\Lambda')$ следует, что

$$\left\| \sum_{j\in A} \lambda_j \boldsymbol{h}_j - \sum_{j\in B} \lambda'_j \boldsymbol{h}_j \right\| > 2\delta. \tag{30}$$

В [36] было показано, что такие коды не существуют. Действительно, в качестве контрпримера положим $A = \{1, 2, \ldots, t\}$, $B = \{1, 2, \ldots, t-1, t+1\}$ и

$$\lambda_t = \lambda'_{t+1} = \lambda < \frac{\delta}{\|\boldsymbol{h}_t\| + \|\boldsymbol{h}_{t+1}\|}.$$

Выберем положительные $\lambda_j = \lambda'_j$ для $j = 1, \ldots, t-1$ такими, что $\lambda + \sum\limits_{j=1}^{t-1} \lambda_j = 1$. Тогда

$$\left\| \sum_{j\in A} \lambda_j \boldsymbol{h}_j - \sum_{j\in B} \lambda'_j \boldsymbol{h}_j \right\| = \|\lambda(\boldsymbol{h}_t - \boldsymbol{h}_{t+1})\| < \delta,$$

и неравенство (30) не выполнено.

В [24] было предложено ограничиться только атакой усреднения и исследовать коды, находящие коалицию целиком в этом случае.

Определение 9. Двоичный код $\mathcal{H}$ называется (t, δ)-*мультимедийным кодом со свойством полного поиска коалиций, устойчивым к атаке усреднения и δ-шуму* ((t, δ)-МППК-кодом), если для любых двух различных подмножеств кода $A, B \subset \mathcal{H}$, таких что $|A|, |B| \leqslant t$, справедливо неравенство

$$\left\| \frac{1}{|A|} \sum_{j\in A} \boldsymbol{h}_j - \frac{1}{|B|} \sum_{j\in B} \boldsymbol{h}_j \right\| > 2\delta. \tag{31}$$

Обозначим через $\mathcal{M}(m, t, \delta)$ максимальную мощность (t, δ)-МППК-кода и определим соответствующую максимальную скорость

$$\mathcal{R}(m, t, \delta) := m^{-1} \log_2 \mathcal{M}(m, t, \delta).$$

Основным результатом [24] стало доказательство существования (t, δ)-мультимедийных кодов со скоростью, отделенной от нуля. А именно: для фиксированных t и δ

$$\liminf_m \mathcal{R}(m, t, \delta) \geqslant \frac{\gamma_t \log_2 e}{t(1 + \gamma_t \log_2 e)} > \frac{\log_2 e}{t(e + \log_2 e)} > \frac{0{,}346}{t}, \tag{32}$$

где $\gamma_t = (1 - t^{-1})^{t-1}$.

Построение таких кодов в [24] было разбито на две подзадачи: первая – построение (t, δ)-МППК-кодов для случая, когда мощность коалиции заранее известна, а вторая – построение кодов, которые позволяют найти мощность коалиции по ложной копии.

Решение первой подзадачи заключается в построении двоичного кода $\mathcal{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\}$, такого что для любых двух различных подмножеств кода $A, B \subset \mathcal{C}$,

таких что $|A| = |B| \leqslant t$, справедливо неравенство

$$\left\| \sum_{\boldsymbol{c} \in A} \boldsymbol{c} - \sum_{\boldsymbol{c}' \in B} \boldsymbol{c}' \right\| > 2\Delta. \tag{33}$$

Действительно, такой код $\mathcal{C}$ позволит однозначно найти всю коалицию при атаке усреднения и δ-шуме, где $\delta = \Delta/t$, если мощность коалиции заранее известна и не превышает t.

Отметим, вслед за [24], что если в определении 9 неравенство

$$\left\| \frac{1}{|A|} \sum_{j \in A} \boldsymbol{h}_j - \frac{1}{|B|} \sum_{j \in B} \boldsymbol{h}_j \right\| > 2\delta$$

заменить на неравенство

$$\left\| \sum_{j \in A} \boldsymbol{h}_j - \sum_{j \in B} \boldsymbol{h}_j \right\| > 2\delta,$$

то получится определение *двоичных* евклидовых дизъюнктивных кодов, тогда как в задаче о евклидовых дизъюнктивных кодах в качестве $\boldsymbol{h}_j$ рассматривались *произвольные* векторы евклидова пространства, см. [61, 62]. Тем самым, рассматриваемая нами подзадача – это задача о двоичных евклидовых дизъюнктивных кодах с одинаковой мощностью коалиций. Будем такие коды сокращенно называть t-ДЕД-кодами. Опишем построение таких кодов, которое навеяно конструкцией [63] (см. также [59, 60, 64]).

Построение t-ДЕД-кодов. Выберем, как мы уже делали при построении t-МППК-кодов, в качестве кода $\widehat{\mathcal{H}}$ столбцы проверочной $(m \times M)$-матрицы линейного двоичного кода V, исправляющего t ошибок. Закодируем слова $\widehat{\boldsymbol{h}}_1, \ldots, \widehat{\boldsymbol{h}}_M \in \widehat{\mathcal{H}}$ двоичным кодом U длины n, с m информационными символами и минимальным кодовым расстоянием d в метрике Хэмминга. Полученные векторы $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M$ и образуют искомый код $\mathcal{H} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M\}$ с $\Delta = \sqrt{d}/2$, см. [24].

Действительно, для двух произвольных различных подмножеств A, B кода $\mathcal{H}$ одинаковой мощности не более t рассмотрим векторы $\boldsymbol{h}^{(A)} = \sum\limits_{\boldsymbol{h} \in A} \boldsymbol{h}$ и $\boldsymbol{h}^{(B)} = \sum\limits_{\boldsymbol{h} \in B} \boldsymbol{h}$. Так как различные суммы *по модулю* 2 из t и менее векторов $\widehat{\boldsymbol{h}}_j$ различны, то $\boldsymbol{h}^{(A)} \bmod 2 \neq \boldsymbol{h}^{(B)} \bmod 2$. Так как векторы $\boldsymbol{h}^{(A)} \bmod 2$ и $\boldsymbol{h}^{(B)} \bmod 2$ принадлежат коду U (в силу линейности кода) и различны, то

$$d_H\big(\boldsymbol{h}^{(A)} \bmod 2, \boldsymbol{h}^{(B)} \bmod 2\big) \geqslant d,$$

где d_H – расстояние Хэмминга, Применив неравенство

$$\|\boldsymbol{a} - \boldsymbol{b}\|^2 \geqslant d_H(\boldsymbol{a} \bmod 2, \boldsymbol{b} \bmod 2),$$

справедливое для любых двух целочисленных векторов $\boldsymbol{a}$ и $\boldsymbol{b}$, получим искомое утверждение.

Перейдем теперь к построению кодов, которые могут найти мощность коалиции. Как ни странно, эта подзадача оказалась несколько сложнее первой подзадачи.

О п р е д е л е н и е 10. Будем говорить, что двоичный код C *определяет мощность коалиции вплоть до t в условиях δ-шума*, если для любых двух его под-

множеств A и B различной мощности не более t справедливо неравенство

$$\left\| \frac{1}{|A|} \sum_{c \in A} \boldsymbol{c} - \frac{1}{|B|} \sum_{c' \in B} \mathbf{c}' \right\| > 2\delta. \tag{34}$$

С помощью такого кода система сформирует итоговый код, приписывая к словам кода $\mathcal{H}$, построенного выше, в качестве "хвостов" слова кода, определяющего мощность. По "хвостам" система найдет мощность коалиции, а затем с помощью кода $\mathcal{H}$ найдет и саму коалицию.

С этой целью в [24] были введены коды, названные авторами слабыми дизъюнктивными кодами. Оказалось, что такие коды исследуются уже довольно давно, видимо, начиная с работы [65], и под разными именами. Наиболее часто они называются *селекторами* (см. [66–68]) или *t-локально тонкими семействами множеств* (см. [69]).

Будем говорить, перефразируя [66–68], что вектор $\boldsymbol{a} \in A \subset B^n$ *выделяется* из множества A, если существует координата i, такая что $a_i = 1$ и $a'_i = 0$ для всех $\boldsymbol{a}' \in A \setminus \{\boldsymbol{a}\}$. Если для двоичного кода C длины n в любом кодовом подмножестве мощности не более t выделяется не менее r векторов, то такой код называется (n, t, r)-селектором. При $r = 1$ получаем определение слабого t-дизъюнктивного кода из [24], или, если заменить двоичные векторы длины n на соответствующие подмножества, то получим определение t-локально тонкого семейства подмножеств множества из n элементов. Отметим, что при $r = t$ получается определение $(t-1)$-дизъюнктивного кода.

В действительности для построения кодов, определяющих мощность коалиции, в [24] использовалось более общее (и сильное) понятие слабого дизъюнктивного кода с исправлением ошибок. Это определение аналогично определению 8.

Определение 11. Двоичный код называется *слабым (t, T)-дизъюнктивным кодом*, если для любого кодового подмножества A, такого что $2 \leqslant |A| \leqslant t$, существует не менее T координат i, таких что $\pi_i(A) = 1$, где $\pi_i(A) := |\{\boldsymbol{a} \in A : \ a_i = 1\}|$.

В качестве примера заметим, что слабый $(2, T)$-дизъюнктивный код – это двоичный код с минимальным кодовым расстоянием в метрике Хэмминга не менее T.

В [24] методом случайного кодирования была доказана следующая нижняя граница для скорости $R(t, T)$ слабых (t, T)-дизъюнктивных кодов при фиксированных t, T:

$$R(t, T) \geqslant \frac{1}{t}\Big(1 - \frac{1}{t}\Big)^{t-1} \log_2 e > \frac{\log_2 e}{et}. \tag{35}$$

Заметим, что эта асимптотическая граница не зависит от фиксированного T.

В [24] был предложен алгоритм линейной от n сложности, который позволяет найти мощность коалиции с помощью произвольного слабого (t, T)-дизъюнктивного кода, если длина шума

$$\|\boldsymbol{e}\| \leqslant \delta = \frac{\sqrt{T}}{2\sqrt{2}} t^{-2}.$$

Естественно выбрать параметры d и T, возникающие при решении первой и второй подзадач, таким образом, чтобы обеспечиваемый уровень шума δ совпадал, что достигается при $T = 2dt^2$. А далее прямые вычисления дают границу (32) (подробнее см. [24]).

§ 6. Заключение

В статье рассмотрены коды, способные полностью обнаружить коалицию пользователей по сделанной ими нелегальной копии мультимедийного контента. Эти коды исследовались как для исходной непрерывной модели, так и для ее дискретной версии. Возникающие коды есть не что иное, как сигнатурные коды для соответствующих каналов множественного доступа, а именно, взвешенного суммирующего канала и A-канала соответственно. Как нам представляется, исследование непрерывной модели оказывается проще в силу аддитивной структуры соответствующего канала множественного доступа. Так, для этой модели известно, что скорость наилучших кодов имеет порядок $t^{-1}\log t$, тогда как для A-канала верхняя и нижняя границы скорости наилучших кодов R_A различаются по порядку в $t^{-1}\log t$ раз (см. (23)). Это различие характерно для многих дискретных моделей каналов множественного доступа, начиная с дизъюнктивного канала и кодов (см. [70, 71]).

Авторы считают своим приятным долгом выразить благодарность И.В. Воробьеву за полезные обсуждения и замечания.

СПИСОК ЛИТЕРАТУРЫ

1. *Trappe W., Wu M., Wang Z.J., Liu K.J.R.* Anti-Collusion Fingerprinting for Multimedia // IEEE Trans. Signal Process. 2003. V. 51. № 4. P. 1069–1087. https://doi.org/10.1109/TSP.2003.809378
2. *Liu K.J.R., Trappe W., Wang Z.J., Wu M., Zhao H.* Multimedia Fingerprinting Forensics for Traitor Tracing. Cairo, Egypt: Hindawi, 2005.
3. *Cheng M., Miao Y.* On Anti-Collusion Codes and Detection Algorithms for Multimedia Fingerprinting // IEEE Trans. Inform. Theory. 2011. V. 57. № 7. P. 4843–4851. https://doi.org/10.1109/TIT.2011.2146130
4. *Cheng M., Ji L., Miao Y.* Separable Codes // IEEE Trans. Inform. Theory. 2012. V. 58. № 3. P. 1791–1803. https://doi.org/10.1109/TIT.2011.2146130
5. *Wagner N.R.* Fingerprinting // Proc. 1983 IEEE Symp. on Security and Privacy. Oakland, CA, USA. Apr. 25–27, 1983. P. 18–22. https://doi.org/10.1109/SP.1983.10018
6. *Blakley G.R., Meadows C., Purdy G.B.* Fingerprinting Long Forgiving Messages // Advances in Cryptology—CRYPTO'85 (Proc. Conf. on the Theory and Application of Cryptographic Techniques. Santa Barbara, CA, USA. Aug. 18–22, 1985). Lect. Notes Comp. Sci. V. 218. Berlin: Springer, 1986. P. 180–189. https://doi.org/10.1007/3-540-39799-X_15
7. *Chor B., Fiat A., Naor M.* Tracing Traitors // Advances in Cryptology—CRYPTO'94 (Proc. 14th Annu. Int. Cryptology Conf. Santa Barbara, CA, USA. Aug. 21–25, 1994). Lect. Notes Comp. Sci. V. 839. Berlin: Springer, 1994. P. 257–270. https://doi.org/10.1007/3-540-48658-5_25
8. *Chor B., Fiat A., Naor M., Pinkas B.* Tracing Traitors // IEEE Trans. Inform. Theory. 2000. V. 46. № 3. P. 893–910. https://doi.org/10.1109/18.841169
9. *Hollmann H.D.L., van Lint J.H., Linnartz J.-P., Tolhuizen L.M.G.M.* On Codes with the Identifiable Parent Property // J. Combin. Theory Ser. A. 1998. V. 82. № 2. P. 121–133. https://doi.org/10.1006/jcta.1997.2851
10. *Barg A., Cohen G., Encheva S., Kabatiansky G., Zémor G.* A Hypergraph Approach to the Identifying Parent Property: The Case of Multiple Parents // SIAM J. Discrete Math. 2001. V. 14. № 3. P. 423–431. https://doi.org/10.1137/S0895480100376848
11. *Boneh D., Shaw J.* Collusion-Secure Fingerprinting for Digital Data // IEEE Trans. Inform. Theory. 1998. V. 44. № 5. P. 1897–1905. https://doi.org/10.1109/18.705568
12. *Barg A., Blakley G.R., Kabatiansky G.A.* Digital Fingerprinting Codes: Problem Statements, Constructions, Identification of Traitors // IEEE Trans. Inform. Theory. 2003. V. 49. № 4. P. 852–865. https://doi.org/10.1109/TIT.2003.809570
13. *Tardos G.* Optimal Probabilistic Fingerprint Codes // J. ACM. 2008. V. 55. № 2. Art. 10 (24 pp.). https://doi.org/10.1145/1346330.1346335

14. *Кабатянский Г.А.* Идентифицирующие коды и их обобщения // Пробл. передачи информ. 2019. Т. 55. № 3. С. 93–105. https://doi.org/10.1134/S0555292319030070
15. *Egorova E., Fernandez M., Kabatiansky G., Lee M.H.* Signature Codes for the A-Channel and Collusion-Secure Multimedia Fingerprinting Codes // Proc. 2016 IEEE Int. Symp. on Information Theory (ISIT'2016). Barcelona, Spain. July 10–15, 2016. P. 3043–3047. https://doi.org/10.1109/ISIT.2016.7541858
16. *Chang S.C., Wolf J.K.* On the T-User M-Frequency Noiseless Multiple-Access Channel with and without Intensity Information // IEEE Trans. Inform. Theory. 1981. V. 27. № 1. P. 41–48. https://doi.org/10.1109/TIT.1981.1056304
17. *Egorova E., Fernandez M., Kabatiansky G., Lee M.H.* Signature Codes for Weighted Noisy Adder Channel, Multimedia Fingerprinting and Compressed Sensing // Des. Codes Cryptogr. 2019. V. 87. № 2–3. P. 455–462. https://doi.org/10.1007/s10623-018-0551-9
18. *Györfi L., Győri S., Laczay B., Ruszinkó M.* Lectures on Multiple Access Channels. Book draft, 2005. Available at http://www.szit.bme.hu/~gyori/AFOSR_05/book.pdf.
19. *Mathys P.* A Class of Codes for T Active Users out of N Multiple-Access Communication System // IEEE Trans. Inform. Theory. 1990. V. 36. № 6. P. 1206–1219. https://doi.org/10.1109/18.59923
20. *Donoho D.L.* Compressed Sensing // IEEE Trans. Inform. Theory. 2006. V. 52. № 4. P. 1289–1306. https://doi.org/10.1109/TIT.2006.871582
21. *Candes E.J., Tao T.* Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? // IEEE Trans. Inform. Theory. 2006. V. 52. № 12. P. 5406–5425. https://doi.org/10.1109/TIT.2006.885507
22. *Кашин Б.С., Темляков В.Н.* Замечание о задаче сжатого измерения // Матем. заметки. 2007. Т. 82. № 6. С. 829–837. https://doi.org/10.4213/mzm4183
23. *Kabatiansky G., Fernandez M., Egorova E.* Multimedia Fingerprinting Codes Resistant against Colluders and Noise // Proc. 8th IEEE Int. Workshop on Information Forensics and Security (WIFS'2016). Abu Dhabi, UAE. Dec. 4–7, 2016. P. 1–5. https://doi.org/10.1109/WIFS.2016.7823904
24. *Егорова Е.Е., Фернандес М., Кабатянский Г.А., Мяо И.* Существование и конструкции мультимедийных кодов, способных находить полную коалицию при атаке усреднения и шуме // Пробл. передачи информ. 2020. Т. 56. № 4. С. 97–108. https://doi.org/10.31857/S0555292320040087
25. *Wolf J.K.* Born Again Group Testing: Multiaccess Communications // IEEE Trans. Inform. Theory. 1985. V. 31. № 2. P. 185–191. https://doi.org/10.1109/TIT.1985.1057026
26. *Rényi A.* On a Problem in Information Theory // Magyar Tud. Akad. Mat. Kutató Int. Közl. 1961. V. 6. P. 505–516.
27. *Ulam S.M.* Adventures of a Mathematician. New York: Scribner, 1976.
28. *Pelc A.* Solution of Ulam's Problem on Searching with a Lie // J. Combin. Theory Ser. A. 1987. V. 44. № 1. P. 129–140. https://doi.org/10.1016/0097-3165(87)90065-3
29. *Kabatiansky G.A., Egorova E.E.* Adversarial Multiple Access Channels and a New Model of Multimedia Fingerprinting Coding // Proc. 2020 IEEE Conf. on Communications and Network Security (CNS'2020). Avignon, France. June 29 – July 1, 2020. P. 1–5. https://doi.org/10.1109/CNS48642.2020.9162248
30. *Сагалович Ю.Л.* Разделяющие системы // Пробл. передачи информ. 1994. Т. 30. № 2. С. 14–35. http://mi.mathnet.ru/ppi228
31. *Cohen G.D., Schaathun H.G.* Asymptotic Overview on Separating Codes // Tech. Rep. № 248. Dept. of Informatics, Univ. of Bergen. Bergen, Norway, 2003. Available at http://www.ii.uib.no/~georg/sci/inf/coding/hyperpdf/cs03rep.pdf.
32. *Blakley G.R.* Safeguarding Cryptographic Keys // Proc. 1979 National Computer Conf.: Int. Workshop on Managing Requirements Knowledge (MARK). New York. June 4–7, 1979. AFIPS Conf. Proceedings, V. 48. Montvale, NJ: AFIPS Press, 1979. P. 313–317. https://doi.org/10.1109/MARK.1979.8817296
33. *Shamir A.* How to Share a Secret // Comm. ACM. 1979. V. 22. № 11. P. 612–613. https://doi.org/10.1145/359168.359176

34. *Егорова Е.Е.* Обобщение IPP-кодов и IPP-систем множеств // Пробл. передачи информ. 2019. Т. 55. № 3. С. 46–59. https://doi.org/10.1134/S0555292319030045
35. *Zhao H.V., Wu M., Wang Z.J., Liu K.J.R.* Forensic Analysis of Nonlinear Collusion Attacks for Multimedia Fingerprinting // IEEE Trans. Image Process. 2005. V. 14. № 5. P. 646–661. https://doi.org/10.1109/TIP.2005.846035
36. *Fan J., Gu Y., Hachimori M., Miao Y.* Signature Codes for Weighted Binary Adder Channel and Multimedia Fingerprinting // IEEE Trans. Inform. Theory. 2021. V. 67. № 1. P. 200–216. https://doi.org/10.1109/TIT.2020.3033445
37. *Djackov A.G.* On a Search Model of False Coins // Topics in Information Theory (Proc. 2nd Colloq. on Information Theory. Keszthely, Hungary. Aug. 25–30, 1975). Colloq. Math. Soc. János Bolyai. V. 16. Amsterdam: Horth Holland, 1977. P. 163–170.
38. *Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А.* Теория кодов, исправляющих ошибки. М.: Связь, 1979.
39. *Bshouty N.H., Mazzawi H.* On Parity Check $(0,1)$-Matrix over $\mathbb{Z}_p$ // Proc. 22nd Annu. ACM–SIAM Symp. on Discrete Algorithms (SODA'11). San Francisco, CA. Jan. 23–25, 2011. P. 1383–1394. https://dl.acm.org/doi/10.5555/2133036.2133142
40. *Kautz W., Singleton R.* Nonrandom Binary Superimposed Codes // IEEE Trans. Inform. Theory. 1964. V. 10. № 4. P. 363–377. https://doi.org/10.1109/TIT.1964.1053689
41. *Friedman A.D., Graham R.L., Ullman J.D.* Universal Single Transition Time Asynchronous State Assignments // IEEE Trans. Comput. 1969. V. 18. № 6. P. 541–547. https://doi.org/10.1109/T-C.1969.222707
42. *Blackburn S.R.* Probabilistic Existence Results for Separable Codes // IEEE Trans. Inform. Theory. 2015. V. 61. № 11. P. 5822–5827. https://doi.org/10.1109/TIT.2015.2473848
43. *Manin Yu.I.* What Is the Maximum Number of Points on a Curve over $\mathbf{F}_2$? // J. Fac. Sci. Univ. Tokyo Sect. IA Math. 1981. V. 28. № 3. P. 715–720.
44. *Randriambololona H.* $(2,1)$-Separating Systems beyond the Probabilistic Bound // Israel J. Math. 2013. V. 195. № 1. P. 171–186. https://doi.org/10.1007/s11856-012-0126-9
45. *Cohen G., Litsyn S., Zémor G.* Binary B_2-Sequences: A New Upper Bound // J. Combin. Theory Ser. A. 2001. V. 94. № 1. P. 152–155. https://doi.org/10.1006/jcta.2000.3127
46. *Erdős P., Frankl P., Füredi Z.* Families of Finite Sets in Which No Set Is Covered by the Union of Two Others // J. Combin. Theory Ser. A. 1982. V. 33. № 2. P. 158–166. https://doi.org/10.1016/0097-3165(82)90004-8
47. *Erdős P., Frankl P., Füredi Z.* Families of Finite Sets in Which No Set Is Covered by the Union of r Others // Israel J. Math. 1985. V. 51. № 1–2. P. 79–89. https://doi.org/10.1007/BF02772959
48. *Воробьев И.В.* Границы скоростей разделяющих кодов // Пробл. передачи информ. 2017. Т. 53. № 1. С. 34–46. http://mi.mathnet.ru/ppi2225
49. *Бассалыго Л.А., Гельфанд С.И., Пинскер М.С.* Простые методы получения нижних границ в теории кодов // Пробл. передачи информ. 1991. Т. 27. № 4. С. 3–8. http://mi.mathnet.ru/ppi576
50. *Дьячков А.Г., Рыков В.В.* Границы длины дизъюнктивных кодов // Пробл. передачи информ. 1982. Т. 18. № 3. С. 7–13. http://mi.mathnet.ru/ppi1232
51. *Guruswami V., Sudan M.* Improved Decoding of Reed–Solomon and Algebraic-Geometry Codes // IEEE Trans. Inform. Theory. 1999. V. 45. № 6. P. 1757–1767. https://doi.org/10.1109/18.782097
52. *Guruswami V.* List Decoding of Error-Correcting Codes (Winning Thesis of the 2002 ACM Doct. Diss. Competition) // Lect. Notes Comp. Sci. V. 3282. Berlin: Springer, 2005.
53. *Форни Д.* Каскадные коды. М.: Мир, 1970.
54. *Alon N.* Explicit Construction of Exponential Sized Families of k-Independent Sets // Discrete Math. 1986. V. 58. № 2. P. 191–193. https://doi.org/10.1016/0012-365X(86)90161-5
55. *Indyk P., Ngo H.Q., Rudra A.* Efficiently Decodable Non-adaptive Group Testing // Proc. 21st Annu. ACM–SIAM Symp. on Discrete Algorithms (SODA'10). Austin, TX. Jan. 17–19, 2010. P. 1126–1142. https://dl.acm.org/doi/10.5555/1873601.1873692

56. *Сагалович Ю.Л.* Верхняя граница мощности кода состояний автомата // Пробл. передачи информ. 1973. Т. 9. № 1. С. 73–83. http://mi.mathnet.ru/ppi884
57. *Сагалович Ю.Л.* Новые верхние границы мощности разделяющих систем // Пробл. передачи информ. 1993. Т. 29. № 2. С. 109–111. http://mi.mathnet.ru/ppi182
58. *Körner J., Simonyi G.* Separating Partition Systems and Locally Different Sequences // SIAM J. Discrete Math. 1988. V. 1. № 3. P. 355–359. https://doi.org/10.1137/0401035
59. *Kabatiansky G., Vlăduţ S., Tavernier C.* On the Doubly Sparse Compressed Sensing Problem // Cryptography and Coding (Proc. 15th IMA Int. Conf. IMACC'2015. Oxford, UK. Dec. 15–17, 2015). Lect. Notes Comp. Sci. V. 9496. Berlin: Springer, 2015. P. 184–189. https://doi.org/10.1007/978-3-319-27239-9_11
60. *Gritsenko V., Kabatiansky G., Lebedev V., Maevskiy A.* Signature Codes for Noisy Multiple Access Adder Channel // Des. Codes Cryptogr. 2017. V. 82. № 1–2. P. 293–299. https://doi.org/10.1007/s10623-016-0228-1
61. *Ericson T., Györfi L.* Superimposed Codes in $\mathbb{R}^n$ // IEEE Trans. Inform. Theory. 1988. V. 34. № 4. P. 877–880.
62. *Füredi Z., Ruszinkó M.* An Improved Upper Bound of the Rate of Euclidean Superimposed Codes // IEEE Trans. Inform. Theory. 1999. V. 45. № 2. P. 799–802. https://doi.org/10.1109/18.749032
63. *Ericson T., Levenshtein V.I.* Superimposed Codes in the Hamming Space // IEEE Trans. Inform. Theory. 1994. V. 40. № 6. P. 1882–1893. https://doi.org/10.1109/18.340463
64. *Влэдуц С.Г., Кабатянский Г.А., Ломаков В.В.* Об исправлении ошибок при искажениях в канале и синдроме // Пробл. передачи информ. 2015. Т. 51. № 2. С. 50–56. http://mi.mathnet.ru/ppi2169
65. *Komlós J., Greenberg A.G.* An Asymptotically Fast Nonadaptive Algorithm for Conflict Resolution in Multiple-Access Channels // IEEE Trans. Inform. Theory. 1985. V. 31. № 2. P. 302–306. https://doi.org/10.1109/TIT.1985.1057020
66. *Clementi A.E.F., Monti A., Silvestri R.* Selective Families, Superimposed Codes, and Broadcasting on Unknown Radio Networks // Proc. 12th Annu. ACM–SIAM Symp. on Discrete Algorithms (SODA'01). Washington, DC, USA. Jan. 7–9, 2001. P. 709–718. https://dl.acm.org/doi/proceedings/10.5555/365411
67. *Chlebus B.S., Kowalski D.R.* Almost Optimal Explicit Selectors // Fundamentals of Computation Theory (Proc. 15th Int. Symp. FCT'2005. Lübeck, Germany. Aug. 17–20, 2005). Lect. Notes Comp. Sci. V. 3623. Berlin: Springer, 2005. P. 270–280. https://doi.org/10.1007/11537311_24
68. *Cicalese F., Vaccaro U.* Superselectors: Efficient Constructions and Applications // Algorithms (Proc. 18th Annu. European Symp. ESA'2010. Liverpool, UK. Sept. 6–8, 2010. Part I). Lect. Notes Comp. Sci. V. 6346. Berlin: Springer, 2010. P. 207–218. https://doi.org/10.1007/978-3-642-15775-2_18
69. *Alon N., Fachini E., Körner J.* Locally Thin Set Families // Combin. Probab. Comput. 2000. V. 9. № 6. P. 481–488. https://doi.org/10.1017/S0963548300004521
70. *Дьячков А.Г., Воробьев И.В., Полянский Н.А., Щукин В.Ю.* Границы скорости дизъюнктивных кодов // Пробл. передачи информ. 2014. Т. 50. № 1. С. 31–63. http://mi.mathnet.ru/ppi2131
71. *Дьячков А.Г., Воробьев И.В., Полянский Н.А., Щукин В.Ю.* Письмо в редакцию // Пробл. передачи информ. 2016. Т. 52. № 2. С. 111. http://mi.mathnet.ru/ppi2208

Егорова Елена Евгеньевна
Кабатянский Григорий Анатольевич
Сколковский институт науки и технологий (Сколтех)
egorovahelene@gmail.com
g.kabatyansky@skoltech.ru

Поступила в редакцию
03.04.2021
После доработки
15.04.2021
Принята к публикации
15.04.2021