УДК 621.391

ЗАЩИТА ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ ТАБЛИЦЫ СТАНДАРТНОГО РАСПОЛОЖЕНИЯ ДЛЯ КОДА

А.И. МИТЮХИН[1], И.И. АСТРОВСКИЙ[2]

[1]Институт информационных технологийБелорусского государственного университета информатики и радиоэлектроники, [2]Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 30 марта 2020

Аннотация. Рассматривается метод защиты информации с использованием разложения группы на смежные классы кода, корректирующего ошибки. Анализируется эффективность перехвата информации в двоичном симметричном канале.

Ключевые слова: кодирование информации, систематическая группа, поле Галуа, смежные классы, мощность кода.

Введение

Одним из алгоритмов декодирования линейного $[n,k]$ - кода над полем Галуа $GF(2)$ является алгоритм, базовой операцией которого является разложение систематической группы G порядка 2^n намножество $C_r = \{C_i\}, i = 1,...,2^r$ смежных классов, где n, k, r соответственно, длина, число информационных ипроверочных символов кода. Декодирование $[n,k]$ – кода сводится к анализу таблицы стандартного расположения для кода [1]. Таблица размером $2^r \times 2^k$ строится на основе операции разложения группы G на смежные классы. Тот факт, что все элементы одного и того же смежного класса имеют один и тот же синдром позволяет упростить процедуру декодирования легальному пользователю системы. Сложность декодирования оценивается только размером или объемом V памяти, необходимым для хранения столбца лидеров смежных классов. Так как столбец имеет размерность $2^r \times n$, то требуемый для декодирования объем памяти определяется как $V = n2^r$ бит.

Как видно, вычислительная сложность декодирования зависит от длины вектора n и соотношения параметров k и r. Если использовать низкоскоростное кодирование, когда $r \square k$, даже для сравнительно небольшой длины n эффективное декодированиепо таблице стандартного расположения для кода на множестве 2^r векторов практически осуществить невозможно. Фактор значительной сложности декодирования при использовании таблицыпредлагается использовать для рассматриваемого метода защиты информации от перехвата нелегальным пользователем.

Теоретические принципы

Пусть имеется блоковыйисточник U^K информации мощностью 2^k информационных векторов вида $\mathbf{u} = (u_1,...,u_k)$. Каждому вектору $\mathbf{u}$ однозначно соответствует кодовое слово линейного кода, записываемое в виде n-мерного вектора $\mathbf{X} = (x_1,...,x_n), x_i \in GF(2)$. Предполагается, что класс и параметры кодов, применяемых в легальном (основном) канале могут с определенной степенью достоверности известны перехватывающей стороне. Имеются также сведения о возможных методах декодирования.В качестве модели канала с защитой информации рассматривается канал, показанный на рис. 1.

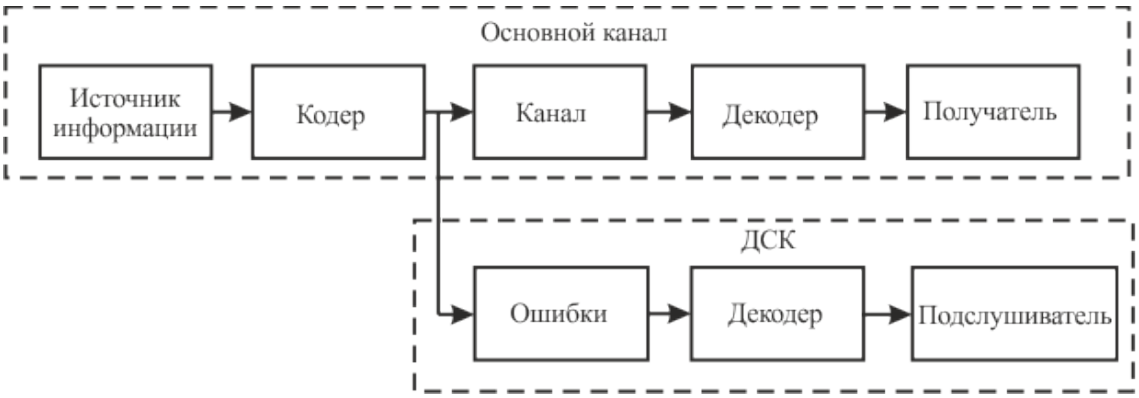

Рис.1. Модель канала с защитой информации

На рис. 1 обозначение ДСК соответствует двоичному симметричному каналу. Декодирование в ДСК сводится к анализу смеси

$$\mathbf{Y} = \mathbf{X} + \mathbf{E}, \qquad (1)$$

где $\mathbf{Y} = (y_1,...,y_n), y_i \in GF(2)$ – вход декодера, $\mathbf{E} = (e_1,...,e_n), e_i \in GF(2)$ – шумовой вектор, препятствующий правильному декодированию в канале подслушивания.

Если передаваемые кодовые слова имеют равные вероятности, оптимальной стратегией правильного приема является анализ и решение на основе принципа максимального правдоподобия. Декодер, анализируя вектор (1), должен решить, какой вектор $\mathbf{X}_i$ из множества $\{\mathbf{X}\}$ передавался. В общем случае, декодирование сводится к нахождению наиболее вероятного вектора ошибок $\mathbf{E}$ для принятого вектора $\mathbf{Y}$. В этом случае декодер работает с минимально возможной ошибкой декодирования. Алгоритм декодирования включает в себя выполнение линейного преобразования для вычисления синдрома

$$\mathbf{S} = \mathbf{Y}\mathbf{H}^T = \mathbf{E}\mathbf{H}^T.$$

В качестве ядра преобразования выступает проверочная матрица $\mathbf{H}$ кода. Далее по $\mathbf{S}$ находится соответствующий смежный класса, вектор ошибок $\mathbf{E}$. Производится оценка вектора $\tilde{\mathbf{X}} = (\mathbf{Y} - \mathbf{E})$, находится информационный вектор $\mathbf{u}$.

Значительную неопределенность получения правильной информации $\mathbf{u}$ по ДСК, можно достичь, если подмножество $\{\mathbf{X}\}$ распределить по всему n-мерному пространству, а не по одному смежному классу согласно определению операции разложениягруппы G на смежные классы. Каждому «правильному» вектору $\mathbf{X}_i \in \{\mathbf{X}\}$ из разрешенного для передачи подмножества группы G следует поставить в соответствие запрещенное подподмножество $\{\mathbf{Y}\}$ (1). Этого можно достичь, если перейти ккоду, ортогональному исходному $[n,k]$-коду. Тогда группа G порядка 2^n раскладывается по подгруппе H порядка 2^r. Любой, искаженный шумом вектор $\mathbf{Y} \neq \mathbf{X}_i$, но находящийся в смежном классе C_i передается как по основному каналу, так и по ДСК.

Для оценки степени защиты информации за счет распределения всех кодовых векторов $\{\mathbf{X}\}$ по множеству $\{C_j\}$, $j = 1,...,2^k$ смежных классов и зашумления вида (1) воспользуемся энтропийным подходом[1]. Надежная защита информации связана с оценкой средней взаимной информация I_u на выходе декодера ДСК.

Пусть первичный источникс алфавитом $U = \{0,1\}$ формирует символы 0 и 1 с вероятностями, соответственно, $p(u_0)$, $p(u_1)$. Источник характеризуется свойством $p(u_0) = p(u_1)$. Тогда, среднее количество информации на слово длиной n определяется как

$$I_u = \frac{k}{n} H(U) - H(U|Y), \tag{2}$$

где $H(U) = \sum_{u_i \in U} p(u_i) \log_2 (p(u_i)$ – энтропия источника, $H(U|Y) = \sum_{u_i \in U} \sum_{y_j \in Y} p(u_i, y_j) \log_2 (p(u_i | y_j)$ – условная энтропия ДСК (потеря информации), $p(u_i, y_j)$ – совместная вероятность символов входа и выхода ДСК, $p(u_i | y_j)$ – апостериорная вероятность символов ДСК.

Для того, чтобы перехват информации не стал возможен, на выходе декодера ДСК должно выполняться условие $I_u = 0$. Это условие можно назвать полной потерей информации в ДСК из-за воздействия на полезную информацию вектора ошибок $\mathbf{E}$. Таким образом, в идеальном случае защиты информации, выражение (2) примет вид

$$\frac{k}{n} H(U) = H(U|Y). \tag{3}$$

Известно [2], условная энтропия $H(U|Y)$ ДСК определяется значениями переходных вероятностейканала $p(0|1) = p(1|0) = p$, т.е. вероятностью ошибок в ДСК. Энтропия ДСК, какэнтропийная функцией Шеннона, определяется формулой [2]

$$H(U|Y) = -[p \log_2 p + (1-p) \log_2 (1-p)]. \tag{4}$$

В рассматриваемом подходе вероятность ошибки ДСК можно связать с весом $\mathrm{wt}(\mathbf{E})$ вектора ошибок (кратностью ошибок t) длиной n.

С учетом равенства (3) получаем выражение

$$k \le n[-p \log_2 p - (1-p) \log_2 (1-p)]. \tag{5}$$

С позиции защиты информации, правая часть (5) определяет энтропию шумовой составляющей на выходе декодера ДСК, когда выполняется условие (3). Выражение (5) позволяет выбрать параметры кода n, r, кратность ошибок t, в зависимости от заранее определяемой степени защиты. Для многих приложений степень защиты определяется длиной n и мощностью кода или количеством 2^{n-r} законов модуляции, используемых в системе. Предварительный расчет необходимых параметров ортогонального кода для модели передачи с защитой информации, рис. 1, покажем на примере.

Пусть $n = 127$, $p = 0{,}1$. Для этого условия вектор ошибок $\mathbf{E}$ имеет вес $\mathrm{wt}(\mathbf{E}) \cong 13$. Используя границу Синглтона [2], оцениваем минимально возможное число проверочных символов. Имеем $r_{\min} = 26$. Заметим, большинство помехоустойчивых кодов имеют намного больше проверочных символов, чем требует граница Синглтона. Используя (4), вычисляем энтропию ДСК. Получаем $H(U|Y) = 0{,}469$. Из выражения (5) следует, что число информационных символов должно быть не больше числа $k \le 59{,}563$. Тогда, число смежных классов $|C_r| \le 2^{59}$. Размерность ортогонального кода, используемого для создания таблицы стандартного расположения, должна быть не больше $k_\perp \le 68$. В каждом смежном классе содержится $M \le 2^{68}$ векторов длиной $n = 127$. Для передачи по ДСК применяется случайное кодирование (1) с использованием случайных векторов $\mathbf{E}$. Количество различных конфигураций векторов $\mathbf{E}$ определяется значением биноминального коэффициента $C_n^{\mathrm{wt}(\mathbf{E})} \cong 1{,}9 \cdot 10^{17}$. Можно утверждать, что векторы $\mathbf{Y}_i = \mathbf{X}_i + \mathbf{E}_i$ также образуют псевдослучайные массивы в каждом смежном классе. Эти массивы двоичных последовательностей не обладает избыточностью. Такое свойство потока символов важно с точки зрения защиты информации от перехвата. Осуществить эффективное однозначное декодирование в канале подслушивателя повектору синдрома $\mathbf{S}$ практически невозможно из-за

значительных вычислительных и временных затрат.Далее следует произвести уточняющие расчеты исходя из конкретного назначения информационной системы, скорости кода, выбранной конструкции кода, необходимости его укорочения или расширения и других особенностей, связанных с прикладным техническим кодированием.

Заключение

Рассмотрен метод защиты информации с использованием теории алгебраических групп. Показано, что использование операцииразложения группы на смежные классы, структурных закономерностей в строении кодов, корректирующих ошибки, позволяет осуществить надежную защиту информации.

PROTECTION OF INFORMATION USING STANDARD LOCATION TABLES FOR CODE

A.I. MITSIUKHIN, I.I. ASTROVSKI

Abstract. The method of information protection using decomposition of the group into adjacent classes of error-correcting code is considered.The efficiency of information interception in a symmetrical binary channel is analyzed.

Keywords: information coding, systematic group, Galois field, group, related classes, code frailty.

Список литературы

1. МитюхинА. И. Прикладная теория информации. Минск, БГУИР. 2018.
2. Mac Willams F.J, Sloane N. J.A. The Theory of Error-Correcting Codes. Oxford, 1977.