Математика и её приложения: ЖИМО. 2008. Выпуск 1 (5). С. 53 – 76.

УДК 512.54

А. А. Толстопятов[1]

Алгоритм кодирования и декодирования поля принадлежности

Ключевые слова: булев полином, сжатие информации.

Построен алгоритм получения коэффициентов булева уравнения, кодирующих поле принадлежности одного буфера по его решениям в базисах Жегалкина и Лагранжа. Построен алгоритм получения кодирующего уравнения, порождающего базиса и кодов булевых полиномов всех буферов файла. Рассмотрены методы решения кодирующего уравнения. Построен алгоритм декодирования поля принадлежности. Обсужден вопрос об адаптации кода поля принадлежности к конкретному файлу.

We construct the algorithm of reception coefficients for boolean equations, which code the belonging field of some buffer by its solutions in Gegalkin's bases and Lagrange's one. We construct also the algorithm of reception the coding equation, generating basis and boolean polynom codes for all buffers of some file. We consider methods of solving the coding equation. We construct the algorithm of decoding of the belonging field. We discuss the question about adaptation of the belonging field code to given file.

Стандартные методы сжатия дискретной информации [1, 3, 5, 17, 18] основаны на повторах в разбиении файла на кортежи. Если различных кортежей достаточно мало, так что каждый из них встретился настолько много раз, что можно набрать достоверную статистику, и частоты появления разных кортежей стабилизируются, то сжатие может быть основано на 1-й теореме Шеннона [16] и алгоритме Хаффмана [15]. Если повторы встречаются всего несколько раз, то сжатие возможно с использованием метода скользящего словаря. Однако, 1-я теорема Шеннона предполагает, что кортежи, на которые разбит файл, независимы друг от друга.

Анализ общих условий сжатия дискретной информации [7] позволил предложить подход к сжатию, основанному на существовании нелинейных зависимостей между буферами, объединяющими, вообще говоря, различное число последовательно следующих кортежей в разбиении файла, если кортежи, входящие в отдельный буфер рассматривать как решения булева уравнения [8]. Поскольку булева переменная $x \in GF(2)$, удовлетворяет условию $x^2 = x$, и при решении булева уравнения кортежи появляются не в том порядке, в котором они расположены в буфере, то код буфера должен включать три поля: поле принадлежности, состоящее из коэффициентов булева полинома, нули которого дают кортежи, принадлежащие данному буферу, поле кратности, определяющее числа повторов кортежей, входящих в буфер, и поле порядка, располагающее кортежи с учетом их

[1] Ивановский государственный университет; E-mail: khash2@mail.ru. Работа выполнена при финансовой поддержке РФФИ (проект 07-07-00155).

кратности в том порядке, в котором они входили в исходный буфер. Алгоритмы построения кода поля кратности рассмотрены в [11, 12], а оценка длины кода этого поля и ее асимптотика — в [9]. Для кода поля порядка соответствующие алгоритмы построены в [10, 14].

В настоящей работе будут рассмотрены алгоритмы построения кода поля принадлежности. Однако, сам это код для отдельного буфера сжатия не дает [13], а дает только перекодировку. Для сжатия файла нужно получить нелинейные зависимости между булевыми полиномами, кодирующими поле принадлежности и, используя эти зависимости [8], включить в код всего файла коэффициенты кодирующего уравнения и коэффициенты булевых полиномов порождающего базиса, а в код отдельного буфера — кортеж из нулей и единиц, показывающий какие из полиномов порождающего базиса нужно подставить вместо переменных в кодирующее уравнение, чтобы получить булев полином поля принадлежности для конкретного буфера [8, 13]. Алгоритмы построения такого кода поля принадлежности также рассмотрены в настоящей работе.

1. Постановка задачи

Введем следующие обозначения:
1) n — число булевых переменных;
2) $x_k \in GF(2)$ — k-я булева переменная $k = 1, 2, \ldots, n$;
3) L — число буферов, на которые разбит файл;
4) m_l — число кортежей в l-ом буфере, $l = 1, 2, \ldots, L$;
5) s_l — число различных кортежей в l-ом буфере, $l = 1, 2, \ldots, L$;
6) $f_l(x_k)$ — булев полином поля принадлежности l-го буфера,
$k = 1, 2, \ldots, n,\ \ l = 1, 2, \ldots, L$.

Остальные обозначения будут вводится по мере необходимости. Если речь идет об одном буфере, то индекс l будет опускаться.

Пусть $\{a_1, a_2, \ldots, a_m\}$ — множество кортежей, входящих в буфер, где $a_l \in GF(2) \times GF(2) \times \cdots \times GF(2) = GF^n(2)$, $l = 1, 2, \ldots, m$ — l-й кортеж. Удалим из этого множества все кортежи, которые встречаются больше одного раза, оставив только по одному из различных кортежей. Такая факторизация превращает исходный набор кортежей в множество

$$\{a_1, a_2, \ldots, a_m\}. \tag{1}$$

Пусть $f(x_k)$ — булев полином. Так как каждый кортеж a_r $(r = 1, 2, \ldots, s)$ есть последовательность булевых переменных $x_1^r, x_2^r, \ldots, x_n^r$, которым присвоены значения 0 или 1, то $f(x_k)$ можно рассматривать как функцию от одного кортежа a_r, причем должно выполняться условие

$$f(a_r) = 0, \quad r = 1, 2, \ldots, s. \tag{2}$$

Первая задача, рассматриваемая в этой работе, заключается в том, чтобы по заданным кортежам a_r найти полином $f(x_k)$. Разумеется, эта задача

решается в зависимости от выбранного базиса. Ниже будут рассмотрены два базиса: базис Жегалкина, в котором полином $f(x_k)$ раскладывается по мономам вида $x_{i_1}x_{i_2}\dots x_{i_s}$, $s = 0, 1, \dots, n-1$, и базис Лагранжа, в котором задача решается технически наиболее просто.

Вторая задача, возникающая при построении кода поля принадлежности, решение которой и дает сжатие, — это поиск нелинейных зависмостей между полиномами $f_l(x_k)$ $(k = 1, 2, \dots, n;\ l = 1, 2, \dots, L)$ и их использование для сокращения кода поля принадлежности. Для решения этой задачи необходимо ввести кодирующий полином

$$F(e_1, e_2, \dots, e_I) \tag{3}$$

от булевых переменных e_i $(i = 1, 2, \dots, I)$ и порождающие булевы полиномы $\varphi_p(x_i)$ $(p = 1, 2, \dots, P)$, удовлетворяющие следующему условию: если вместо e_i для l-го буфера взять набор $\varphi_{i_1}, \varphi_{i_2}, \dots, \varphi_{i_I}$ (среди этих порождающих полиномов могут быть и одинаковые), то

$$F(e_i^l) = f_l, \tag{4}$$

где через e_i^l обозначен набор этих подмножеств для l-го буфера.

Уравнение (4) назовем *кодирующим*. Вторая задача будет решена, если по заданному набору булевых полиномов f_l $(l = 1, 2, \dots, L)$ оказывается возможным построить 2^I коэффициентов многочлена (3), порождающие полиномы $\varphi_p(x_i)$ $(p = 1, 2, \dots, P)$ и номера порождающих полиномов для каждого буфера e_i^l $(i = 1, 2, \dots, I;\ l = 1, 2, \dots, L)$.

Третья задача, рассматриваемая в этой работе, — это задача об упрощении решения кодирующего уравнения. Принципиально, как показано в [8] и обсуждено в [13], решение этой задачи сводится к решению булева уравнения, которое всегда может быть сделано последовательным присвоением всем переменным значений 0 и 1. Задача заключается в том, чтобы сократить этот перебор.

Четвертая задача, которая рассматривается в этой работе, — это задача об адаптации кода поля принадлежности к файлу. Такая задача может быть поставлена и обсуждена, т. к. существуют разные базисы, в которых могут записываться булевы полиномы, и можно попытаться преобразовать такой базис к заданному набору булевых полиномов.

2. Базис Жегалкина

Пусть задано множество кортежей (1) и требуется найти коэффициенты полинома f, удовлетворяющего условию (2), причем полином $f(x_1, \dots, x_n)$ зависит от n булевых переменных. Базис Жегалкина $\psi_1, \psi_2, \dots, \psi_{2^n}$ — это базис, построенный из мономов: $1, x_1, x_2, \dots, x_n, x_1x_2, x_1x_3, \dots, x_{n-1}x_n,\ x_1x_2x_3, \dots, x_{n-2}x_{n-1}x_n,\ \dots,\ x_1x_2x_3\cdots x_{n-1}x_n$.

Всего таких полиномов 2^n. Тогда полином $f(x_1, \ldots, x_n)$ в базисе Жегалкина запишется следующим образом

$$\begin{aligned} &f(x_1, x_2, \ldots, x_n) = C_0 + C_1X_1 + C_2X_2 + \cdots + C_nX_n + \\ &+C_{12}X_1X_2 + C_{13}X_1X_3 + \cdots + C_{n-1,n}X_{n-1}X_n + \\ &+C_{123}X_1X_2X_3 + C_{124}X_1X_2X_4 + \cdots + C_{n-2,n-1,n}X_{n-2}X_{n-1}X_n + \\ &\ldots \\ &+C_{12\ldots(n-1)n}X_1X_2\cdots X_{n-2}X_{n-1}X_n, \end{aligned} \tag{5}$$

причем коэффициенты в записи (5) распадаются на $n+1$ групп в зависимости от числа сомножителей k в соответстующих мономах, и каждая группа содержит $C_n^k = \dfrac{n!}{k!(n-k)!}$ коэффициентов, где $k = 0, 1, \ldots, n$. Всего коэффицентов у полинома (5) будет

$$\sum_{k=0}^{n} C_n^k = 2^n. \tag{6}$$

Чтобы найти коэффициенты $C_0, C_1, \ldots, C_n, C_{12}, \ldots, C_{12\ldots(n-1)n}$, составим список всех возможных кортежей a_l. Этих кортежей будет 2^n:

$$\begin{array}{ll} 0. & 00\ldots00 \\ 1. & 00\ldots01 \\ 2. & 00\ldots10 \\ \ldots & \ldots \\ 2^n-1. & 11\ldots11 \end{array} \tag{7}$$

Последовательно подставляя (7) в (5), получим систему из 2^n линейных уравнений для определения $C_0, C_1, \ldots, C_n, C_{12}, \ldots, C_{12\ldots(n-1)n}$, если в правые части этих уравнений подставить 0, когда соответствующий кортеж входит в буфер, и 1, когда он не входит. Для доказательства того, что эта система уравнений имеет единственное решение, а значит, любому набору кортежей, входящих в буфер, соответствует один и только один булев полином, корнями которого являются эти кортежи, удобно подругому упорядочить все кортежи (7). А именно, возьмем первый кортеж, в котором 0 единиц, далее n кортежей, в которых по одной единице, $\ldots$, C_n^k кортежей в которых по k единиц и т. д. до одного кортежа, в котором n единиц:

$$\begin{array}{ll} 0. & 00\ldots00 \\ 1. & 10\ldots00 \\ 2. & 01\ldots00 \\ \ldots & \ldots \\ n-1. & 00\ldots10 \\ n. & 00\ldots01 \\ \ldots & \ldots \\ 2^n-1. & 11\ldots11 \end{array} \tag{8}$$

Если ввести обозначения

$$d = (d_0, \ldots, d_{2^n-1}) = (C_0, C_1, \ldots, C_n, C_{12}, \ldots, C_{12\ldots(n-1)n}), \quad (9)$$

$$b = (b_0, \ldots, b_{2^n}), \quad (10)$$

где $b_k = 0$, когда кортеж с номером k входит в буфер, и $b_k = 1$, когда не входит, то систему для определения коэффициентов булева полинома f можно записать в виде

$$Ad = b, \quad (11)$$

где A — матрица $2^n \times 2^n$, строки которой составлены из коэффициентов при $C_0, C_1, \ldots, C_n, C_{12}, \ldots, C_{12\ldots(n-1)n}$, получающихся при подстановке в (5) кортежей из (8). Нетрудно увидеть, что при упорядочении кортежей (8) выше главной диагонали в матрице A стоят одни нули, а при последовательном вычеркивании столбца и строки, в которой стоит единственная единица, в следующей строке опять будет стоять единственная единица. Это свойство матрицы A можно проиллюстрировать примером для $n = 3$:

№	кортежи	C_0	C_1	C_2	C_3	C_1C_2	C_1C_3	C_2C_3	$C_1C_2C_3$
0	000	1	0	0	0	0	0	0	0
1	100	1	1	0	0	0	0	0	0
2	010	1	0	1	0	0	0	0	0
3	001	1	0	0	1	0	0	0	0
4	110	1	1	1	0	1	0	0	0
5	101	1	1	0	0	0	1	0	0
6	011	1	0	1	0	0	0	1	0
7	111	1	1	1	1	1	1	1	1

Из указанного выше свойства матрицы A следует, что

$$\det A = 1, \quad (12)$$

а значит система (11) — крамеровская, и, следовательно, имеет единственное решение.

Базис Жегалкина неудобен тем, что для получения коэффициентов полинома $f(x_1, \ldots, x_n)$ требуется решать систему 2^n линейных уравнений с 2^n переменными. Это решение облегчается тем, что при упорядочении этой системы согласно (8) матрица A имеет гауссовский вид; поэтому из 1-го уравнения можно найти C_0, из 2-го — C_1 и т. д., а из 2^n-го — $C_{12\ldots n}$.

3. Базис Лагранжа

Базис Лагранжа [6] — это базис главных идеалов в булевой алгебре [2]. В отличие от базиса Жегалкина для получения коэффициентов полинома $f(x_1, \ldots, x_n)$ в базисе Лагранжа не требуется решения никаких

уравнений. Пусть $x \in GF(2)$ — булева переменная. Определим полиномы Лагранжа от этой переменной следующим образом:

$$L_0(x) = x + 1;\ L_1(x) = x. \tag{13}$$

Нетрудно увидеть, что

$$L_0^2 = L_0;\ L_0 L_1 = L_1 L_0 = 0;\ L_1^2 = L_1; \tag{14}$$

поэтому $L_0,\ L_1$ — главные идеалы в кольце булевых полиномов. Любой полином $f(x)$ может быть разложен по $L_0(x),\ L_1(x)$:

$$f(x) = \sum_{j=0}^{1} C_j L_j(x), \tag{15}$$

где $C_j \in GF(2)$. Чтобы по заданному $f(x)$ найти C_j, перепишем (14) в виде

$$L_i L_j = \delta_{ij} L_j,\ i, j = 0, 1, \tag{16}$$

причем в (16) в правой части нет суммирования по j. Тогда из (15) и (16) получим

$$f(x)L_i(x) = \sum_{j=0}^{1} C_j L_j(x) L_i(x) = \sum_{j=0}^{1} C_j \delta_{ij} L_j(x) = C_i L_i(x), \tag{17}$$

т. е. $f(x)L_i(x) = 0 \Rightarrow C_i = 0$ или $f(x)L_i(x) = L_i(x) \Rightarrow C_i = 1$.

Понятие полиномов Лагранжа легко обобщается на случай произвольного числа булевых переменных. Пусть x_i $(i = 1, 2, \ldots, n)$ — булевы переменные. Тогда полином Лагранжа $L_j(x_i)$ есть

$$L_j(x_i) = \prod_{k=1}^{n} L_{j_k}(x_k), \tag{18}$$

где j_k — коэффициент в двоичной записи числа j:

$$j = \sum_{k=1}^{n} j_k 2^{k-1}. \tag{19}$$

Нетрудно увидеть, что свойство (16) сохраняется:

$$L_k(x_i) L_m(x_i) = \delta_{km} L_m(x_i), \tag{20}$$

где $i = 1, 2, \ldots, n;\ k = 0, 1, \ldots, 2^{n-1};\ m = 0, 1, \ldots, 2^{n-1}$. Действительно, если $k \neq n$, то в двоичном разложении чисел k и m хотя бы для одной степени 2^{k-1} коэффициенты различаются, т. е. для этого p-го разряда в

(18) для $L_k(x_i)$ будут стоять $L_{j_p}(x_p)$, а для m будет стоять $L_{j_p+1}(x_p)$, произведение которых вследствие (14) равно 0. Если же $k = m$, то вследствие (14) будем иметь

$$L_k(x_i)L_k(x_i) = L_k(x_i). \tag{21}$$

Таким образом, и для n булевых переменных $x_i,\ i = 1, \ldots, n$, полиномы Лагранжа образуют базис главных идеалов. Так как существуют 2^n различных способов выбрать $L_{j_k}(x_k)$ в (18), то разных полиномов Лагранжа тоже существует 2^n, а значит в (18) $j = 0, 1, \ldots, 2^{n-1}$. Любой булев полином $f(x)$, $i = 1, \ldots, n$, может быть разложен по полиномам Лагранжа:

$$f(x_i) = \sum_{j=0}^{2^n-1} C_j L_j(x_i), \tag{22}$$

причем коэффициенты C_j в силу (21) могут быть найдены из соотношения

$$f(x_i)L_m(x_i) = C_m L_m(x_i), \tag{23}$$

где $C_m = 0$, если $f(x_i)L_m(x_i) = 0$, и $C_m = 1$, если $f(x_i)L_m(x_i) = L_m(x_i)$.

Отметим два свойства полиномов Лагранжа. Первое:

$$\sum_{j=0}^{2^n-1} L_j(x_i) = 1; \tag{24}$$

второе: уравнение

$$L_j(x_i) = 0, \tag{25}$$

где $i = 1, \ldots, n,\ j = 0, 1, \ldots, 2^{n-1}$, для любого j имеет в точности 2^{n-1} решений.

Переход от базиса Лагранжа к базису Жегалкина делается путем раскрытия скобок в соотношении

$$f(x) = \sum_{j=0}^{2^n-1} C_j \prod_{k=1}^{n} L_{j_k}(x_i), \tag{26}$$

где j_k определяется по j согласно (19). Обратный переход от базиса Жегалкина к базису Лагранжа делается, согласно (17), умножением $f(x)$ на полиномы Лагранжа.

Построить булев полином $f(x_1, \ldots, x_n)$ по его корням $a_l,\ 0 \le l \le 2^n$, в базисе Лагранжа можно следующим образом.

1. Любой корень a_l — т. е. последовательность n нулей и единиц, рассматриваем как натуральное число $j,\ 0 \le j \le 2^n - 1,\ 0 \le a_l \le 2^n$.

2. В разложении (22) полагаем $C_j = 0$, если j есть среди a_l, и $C_j = 1$, если его нет.

3. Если $l < 2^{n-1}$, то, используя (24), можно сократить вычисление $f(x)$, положив $C_j = 1$, если j есть среди a_l, и $C_j = 0$, если его нет, и добавив к полученному полиному 1.

Докажем эти правила. Сначала построим полином, который имеет единственный корень $a = (a_1, a_2, \ldots, a_n)$, $a_i \in GF(2)$, $i = 1, 2, \ldots, n$. Рассмотрим уравнение

$$f(x_1, \ldots, x_n) = \sum_{j=0}^{a-1} L_j(x_i) + \sum_{j=a+1}^{2^n-1} L_j(x_i) = 0, \tag{27}$$

построенное в соответствии с правилом 2, где a_i определяются из

$$a = \sum_{i=1}^{n} a_i \cdot 2^{i-1}. \tag{28}$$

Используя (24), можно записать (27) в виде

$$L_a(x_j) + 1 = 0, \tag{29}$$

или, согласно (18)

$$\prod_{k=1}^{n} L_{a_k}(x_k) + 1 = 0. \tag{30}$$

Уравнение (3) превратится в тождество тогда и только тогда, когда для любого k ($k = 1, 2, \ldots, n$)

$$L_{a_k}(x_k) = 1; \tag{31}$$

но, согласно (13):

$$L_{a_k}(x_k) = a_k + x_k + 1. \tag{32}$$

Тогда из (31) и (32) получим

$$a_k + x_k = 0. \tag{33}$$

Но уравнение (33) имеет единственное решение

$$x_k = a_k. \tag{34}$$

Тем самым показано, что если булев полином имеет единственный корень, то этот полином однозначно определяется правилом 2.

Пусть $a_1, a_2, \ldots, a_l$ — корни полинома f; рассматривая согласно правилу 1 a_k ($k = 1, 2, \ldots, l$) как натуральные числа, без ограничения общности можно считать, что $a_1 < a_2 < \cdots < a_l$. Как показано выше, полином $f_1 = L_{a_1} + 1$ будет иметь единственный корень a_1, полином $f_2 = L_{a_2} + 1$

— единственный корень a_2, и т. д., $f_l = L_{a_l} + 1$ — единственный корень a_l. Тогда уравнение

$$f = \prod_{k=1}^{l} f_k = \prod_{k=1}^{l} (L_{a_k} + 1) = 0 \tag{35}$$

будет иметь l корней a_1, a_2, ..., a_l. Раскрывая скобки в левой части (35) и пользуясь (20), запишем (35) следующим образом:

$$\sum_{k=1}^{l} L_{a_k} + 1 = 0. \tag{36}$$

Используя (24), разложим 1 в (36) по полиномам Лагранжа и выделим в этом разложении члены, присутствующие в сумме в левой части (36). Тогда будем иметь:

$$\begin{aligned} 1 + \sum_{k=0}^{l} L_{a_k} &= \sum_{j=0}^{2^n-1} L_j + \sum_{k=0}^{l} L_{a_k} = \\ &= \sum_{j=0}^{a_1-1} L_j + \sum_{j=a_1+1}^{a_2-1} L_j + \sum_{j=a_2+1}^{a_3-1} L_j + \cdots + \sum_{j=a_l+1}^{2^n-1} L_j + \\ &+ \sum_{k=1}^{n} (L_{a_k} + L_{a_k}) = 0. \end{aligned} \tag{37}$$

Так как для любого булева полинома f справедливо равенство $f + f = 0$, то последняя сумма в (37) изчезает:

$$f = \sum_{j=0}^{a_1-1} L_j + \sum_{j=a_1+1}^{a_2-1} L_j + \cdots + \sum_{j=a_l+1}^{2^n-1} L_j = 0. \tag{38}$$

Но полином в левой части (38), имеющей корнями a_1, a_2, ..., a_l, построен в соответствии с правилом 2.

4. Кодирующее уравнение

Код отдельного l-го буфера состоит из трех полей [8]:

Поле принадлежности	поле кратности	поле порядка
$C_0^l C_1^l \ldots C_{2^n-1}^l$	$N_1^l \quad N_2^l$	N

Рис. 1. Структура кода отдельного буфера

где C_j^l — коэффициенты разложения булева полинома $f_l(x_i)$ $(i = 1, \ldots, n)$ по базису Лагранжа

$$f_l(x_i) = \sum_{j=0}^{2^n-1} C_j^l L_j(x_i), \tag{39}$$

N_1^l — номер стандартной формы в таблице стандартных форм поля кратности [12], N_2^l — номер перестановки в таблице перестановок заданной стандартной формы [12], N — номер перестановки в таблице поля порядка [10]. Чтобы сократить длину кода поля принадлежности необходимо ввести еще одно поле, но не для каждого буфера, а для всего файла. Назовем это поле общим. Тогда структура кода всего файла будет выглядеть так:

Общее поле	1	2	...	L

Рис. 2. Структура кода файла, разбитого на L буферов

Для построения кода общего поля введем кортеж $(e_1, e_2, \ldots, e_I)$ из I булевых полиномов и кодирующий полином $F = F(e_1, e_2, \ldots, e_I)$, который удобно задать коэффициентами разложения F по полиномам Лагранжа $L_j^e,\ j = 0, 1, \ldots, 2^I - 1$, построенным из булевых полиномов e_i:

$$L_j^e = L_j^e(e_i),\ i = 1, \ldots, I,\ j = 0, \ldots, 2^I - 1, \tag{40}$$

т. е.

$$F(e_i) = \sum_{j=0}^{2^I-1} C_j^e L_j^e(e_i). \tag{41}$$

Полиномы порождающего базиса $\varphi_p,\ p = 1, 2, \ldots, P$, зависят от булевых переменных $x_i,\ i = 1, 2, \ldots, I$. Полином F даст кодирующее уравнение, если потребовать, чтобы при замене e_k для l-го буфера на один из φ_p,

$$e_k^l = \varphi_{l_k}, \tag{42}$$

кодирующий полином давал полином поля принадлежности f_l:

$$F(e_k^l) = f_l(x_i),\ k = 1, 2, \ldots, I;\ i = 1, 2, \ldots, n. \tag{43}$$

Тогда структура общего поля примет вид:

$C_0^l C_1^l \ldots C_{2^I-1}^l$	φ_1 φ_2 ... φ_P
коэффициенты кодирующего полинома	коэффициенты полиномов порождающего базиса

Рис. 3. Структура кода общего поля

За счет введения общего поля изменится код поля принадлежности отдельного буфера. А именно, вместо коэффициентов C_j^e из разложения (39) будем иметь номера полиномов порождающего базиса, которые надо взять в качестве e_k^l, так что код поля принадлежности приобретет вид:

e_1^l	e_2^l	...	e_I^l
l_1	l_2	...	l_I

Рис. 4. Структура кода поля принадлежности после введения общего поля

где $l_1, l_2, \ldots, l_I \in \{1, 2, \ldots, P\}$.

Таким образом, для построения кода поля принадлежности необходимо по заданным $f_l(x_i)$ $(i = 1, 2, \ldots, n;\ l = 1, 2, \ldots, L)$ найти коэффициенты кодирующего полинома C_j^l $(j = 0, 1, \ldots, 2^I - 1)$ из (41), полиномы порождающего базиса $\varphi_p(x_i)$ $(p = 1, 2, \ldots, P)$ которые удобно задать коэффициентами C_{pj} $(p = 1, \ldots, P;\ j = 0, 1, \ldots, 2^n - 1)$ в разложении их по полиномам Лагранжа,

$$\varphi_p(x_i) = \sum_{j=0}^{2^n-1} C_{pj} L_j(x_i), \tag{44}$$

и номера полиномов порождающего базиса для l-го буфера $l_1, l_2, \ldots, l_I$, чтобы выполнилось условие (43). Как любая система булевых уравнений, система (43) может быть сведена к одному булеву уравнению [8, 13]. Однако, такое сведение не всегда оправдано, т. к. система (43) содержит слишком много переменных. Поэтому имеет смысл рассмотреть методы решения системы кодирующих уравнений (43), специально разработанные для этой системы.

5. Методы решения кодирующего уравнения

Структура кодирующего уравнения зависит от соотношения параметров, характеризующих разбиение файла длиной N_Φ на кортежи и объединение этих кортежей в буферы. Для поля принадлежности эти параметры есть:

1) n — длина кортежа,

2) L — число буферов,

3) P — число порождающих булевых полиномов,

4) I — число булевых переменных $e_1, e_2, \ldots, e_I$ в кодирующем полиноме $F(e_i)$.

Длина кода поля принадлежности не зависит от числа кортежей в l-м буфере m_l, поэтому любые объединения кортежей в буферы не скажутся на длине кода поля принадлежности, хотя от него зависят длины кодов полей кратности и порядка. Ограничение значений m_l определяется только условием

$$n \sum_{l=1}^{L} m_l = N_\Phi. \tag{45}$$

Параметры n, L, P, I связаны друг с другом неравенством, требующим, чтобы суммарная длина всех L полей принадлежности до введения

общего поля была больше длины общего поля плюс суммарная длина кодов всех L полей принадлежности после введения этого поля. Суммарная длина кодов L полей принадлежности есть:

1) 2^n — длина кода поля принадлежности до введения общего поля;

2) $2^n L$ — суммарная длина кода L полей принадлежности до введения общего поля;

3) 2^I — длина кода коэффициентов кодирующего полинома $F(e_i)$, $i = 1, 2, \ldots, I$;

4) 2^n — длина кода коэффициентов порождающего полинома $\varphi(x_i)$, $i = 1, 2, \ldots, n$;

5) $2^n P$ — суммарная длина кодов коэффициентов всех P порождающих полиномов $\varphi_p(x_i),\ p = 1, 2, \ldots, P;\ i = 1, 2, \ldots, n$;

6) $\log_2(P)$ — длина кода подстановки $e_i \to \varphi_{p_i},\ i = 1, 2, \ldots, I$;

7) $I \log_2(P)$ — суммарная длина кодов кортежей $l_1, l_2, \ldots, l_I$ при подстановке $e_i^l \to \varphi_{p_i}^l$.

Поэтому, чтобы выделение общего поля сокращало суммарный код L полей принадлежности, должно выполняться неравенство:

$$2^I + 2^n P + LI \log_2 P < 2^n L. \tag{46}$$

Если ввести две функции от I:

$$\begin{aligned} f_1(I) &= 2^I, \\ f_2(I) &= 2^n(L - P) - (L \log_2 P)I \end{aligned} \tag{47}$$

и их разность

$$F(i) = f_2(I) - f_1(I), \tag{48}$$

то неравенство (46) примет вид

$$F(I) > 0. \tag{49}$$

Если учесть, что $I \geq 1$, то неравенство (49) будет выполняться, если $1 \leq I < I_{\max}$, где $I_{\max}$ определяется из уравнения

$$F(I_{\max}) = 0. \tag{50}$$

Условие существования решения уравнения (50) есть

$$2^n(L - P) > 1, \tag{51}$$

откуда получаем первое ограничение на параметры разбиения файла:

$$P < L - 2^{-n}. \tag{52}$$

Далее, I удовлетворяет неравенствам

$$1 < I < I_{\max} < \frac{2^n(L - P)}{L \log_2(P)}. \tag{53}$$

Чтобы найти $I_{\max}$ запишем (50), подставив в него (47) и (48):

$$2^{I_{\max}} = 2^n(L-P)\left[1 - \frac{L\log_2 P}{2^n(L-P)}\right]. \tag{54}$$

Логарифмируя (54), найдем, что

$$I_{\max} = n + \log_2(L-P) + \log_2\left[1 - \frac{L\log_2 P}{2^n(L-P)}\right]. \tag{55}$$

Вследствие (53) выполнено неравенство

$$\frac{L\log_2 P}{2^n(L-P)} I_{\max} < 1. \tag{56}$$

Поэтому, раскладывая в (55) логарифм в степенной ряд, ограничиваясь линейными членами и решая полученные уравнения относительно $I_{\max}$, найдем, что

$$I_{\max} = \frac{2^n(L-P)[n + \log_2(L-P)]}{2^n(L-P) + L\log_2 e \log_2 P}. \tag{57}$$

Поскольку из (52) следует, что выполняется как

$$\frac{P}{L} < 1 - \frac{1}{2^n L} < 1, \tag{58}$$

так и

$$\frac{1}{2^n L} < 1 - \frac{P}{L} < 1, \tag{59}$$

то, записывая $\log_2 P$ в виде

$$\log_2 P = \log_2 L - \log_2\left[1 - \left(1 - \frac{P}{L}\right)\right], \tag{60}$$

раскладывая второй логарифм в правой части (60) в ряд по степеням $(1 - \frac{P}{L}) < 1$ и сохраняя только линейный член, получим

$$\log_2 P = \log_2 \frac{L}{e} + \frac{P}{L}\log_2 e. \tag{61}$$

Для методов решения кодирующего уравнения самым важным является соотношение между I и P. Это соотношения можно получить из (53), используя неравенство $I < I_{\max}$ и (57). Однако, в этом случае формулы получаются громоздкими по сравнению со случаем, когда используется более слабое неравенство, вытекающее из (53):

$$I < \frac{2^n(L-P)}{L\log_2 p}. \tag{62}$$

Поскольку качественно в обоих этих случаях выводы совпадают, то в дальнейшем будем использовать (62). Подставляя (61) в (62) получим:

$$I < \frac{2^n}{\log_2 \frac{L}{p}} \cdot \frac{1 - \frac{P}{L}}{1 + \frac{\log_2 e}{\log_2 \frac{L}{e}} \cdot \frac{P}{L}}. \tag{63}$$

Разлагая (63) в ряд по $\frac{P}{L} < 1$, сохраняя линейный член и учитывая, что $L >> e$, приводим (63) к виду:

$$\frac{\log_2 L}{2^n} I + \frac{P}{L} < 1. \tag{64}$$

Из (64) следует, что условие сжатия (46) совместимо и со случаем $I > P$ и со случаем $I < P$.

Еще одно полезное неравенство можно получить, переписав (46) в виде:

$$2^I + 2^n P < (2^n - I \log_2 P) L. \tag{65}$$

Из (65) сразу же следует, что

$$I \cdot \log_2 P < 2^n. \tag{66}$$

Поскольку разбиение файла на буферы начинается с разбиения его на кортежи равной длины n, то неравенство (66) простым способом ограничивает допустимые значения I и P, если известно n.

Наконец, найдем соотношение между L и I, которое существенно при решении кодирующего уравнения. Из (65) имеем:

$$L > \frac{2^I + 2^n P}{2^n - I \log_2 P}. \tag{67}$$

Из (67) будем иметь, что

$$L > \frac{2^I + 2^n P}{2^n - I \log_2 P} > \frac{2^I}{2^n} = 2^{I-n}. \tag{68}$$

Логарифмируя (68) получим искомое неравенство:

$$I < n + \log_2 L. \tag{69}$$

Для того, чтобы найти решения кодирующего уравнения (43), необходимо по заданным полиномам $f_l(x_i)$, $i = 1, \ldots, n$, т. е. по заданным коэффициентам C_j^l ($l = 1, \ldots, L$; $j = 0, \ldots, 2^n - 1$) в разложении (39) найти следующие переменные:
1) C_{pj} в разложении (44), т. е. порождающие полиномы φ_p ($p = 1, \ldots, P$);
2) C_j^e в разложении (41), т. е. коэффициенты порождающего полинома $F(e_i)$ ($i = 1, \ldots, I$);

3) подстановки (42) вида

$$\{e_1^l e_2^l \dots e_I^l\} \to \{\varphi_{l_1} \varphi_{l_2} \dots \varphi_{l_I}\}, \tag{70}$$

причем при $I \le P$ все l_i $(i = 1, \dots, I)$ могут, но не обязаны быть разными, а при $I > P$ среди l_i $(i = 1, \dots, I)$ обязательно есть одинаковые.

Идея решения кодирующего уравнения (43) заключается в том, чтобы отделить задачу о существовании базиса от задачи о существовании кодирующего уравнения. Это может быть сделано потому, что хотя в разложении (39) полиномы $f_l(x_i)$ рассматриваются как булевы полиномы от переменных x_i, $i = 1, \dots, n$, но все они должны быть выражены через порождающий базис φ_p, $p = 1, \dots, P$, т. е.

$$f_l = f(\varphi_p),\ l = 1, \dots, L;\ p = 1, \dots, P. \tag{71}$$

Для получения уравнений, определяющих порождающие полиномы φ_p, запишем (71) в виде:

$$f_l = \sum_{j=0}^{2^n-1} C_j^l L_j(x) = \sum_{i=1}^{2^P-1} a_i^l L_i(\varphi), \tag{72}$$

где полиномы f_l разложены по полиномам Лагранжа, построенным из φ_p согласно

$$L_i(\varphi) = \prod_{p=1}^{P} L_{i_p}(\varphi_p), \tag{73}$$

а i_p — коэффициенты в двоичном представлении числа i:

$$i = \sum_{p=1}^{P} I_p 2^{p-1}\ . \tag{74}$$

Уравнение (72) позволяет найти по заданным C_j^l не только a_i^l, но и C_{pj} из разложения (44), т. е. порождающие полиномы φ_p. Для этого разложим $L_j(\varphi)$, $i = 0, \dots, 2^P - 1$, по $L_j(x)$, $j = 0, \dots, 2^n - 1$. Согласно (73) будем иметь

$$L_i(\varphi) = \prod_{p=1}^{P} L_{i_p}(\varphi_p) = \prod_{p=1}^{P} (1 + i_p + \varphi_p) = \prod_{p=1}^{P} (1 + i_p + \sum_{j=0}^{2^n-1} C_{pj} L_j(x)). \tag{75}$$

Так как $i_p \in \{0, 1\}$, то можно записать:

$$1 = \sum_{j=0}^{2^n-1} L_j(x), \quad i_p = \sum_{j=0}^{2^n-1} i_p L_j(x). \tag{76}$$

Поставляя (76) в (75), получим, что

$$L_i(\varphi) = \prod_{p=1}^{P} \sum_{j=0}^{2^n-1} (1 + i_p + C_{pj}) L_j(x). \tag{77}$$

Так как произведения разных полиномов Лагранжа равны 0, а одинаковых — самому полиному Лагранжа, то в (77) можно поменять местами произведение и сумму:

$$L_i(\varphi) = \sum_{j=0}^{2^n-1} \left[\prod_{p=1}^{P} (1 + i_p + C_{pj}) \right] L_j(x). \tag{78}$$

Подставляя (78) в (72), меняя порядок суммирования по $i = 0, \ldots, 2^P - 1$ и $j = 0, \ldots, 2^n - 1$ и приравнивая коэффициенты при одинаковых полиномах Лагранжа, получим систему булевых уравнений для определения a_i^l и c_{pj} по заданным C_j^l:

$$\sum_{i=0}^{2^P-1} \left[a_i^l \prod_{p=1}^{P} (1 + i_p + C_{pj}) \right] + C_j^l = 0. \tag{79}$$

Система уравнений (79) содержит $L \cdot 2^P$ неизвестных a_i^l, $P \cdot 2^n$ неизвестных C_{pj} и состоит из $L \cdot 2^n$ уравнений. Её можно записать как одно булево уравнение, взяв дизъюнкцию по индексам j и l:

$$\bigvee_{j=0}^{2^n-1} \bigvee_{l=1}^{L} \left\{ \sum_{i=0}^{2^P-1} \left[a_i^l \prod_{p=1}^{P} (1 + i_p + C_{pj}) \right] + C_j^l \right\} = 0. \tag{80}$$

Таким образом, показано, что при фиксированных n и L уравнение (80) определяет все возможные базисы порождающих полиномов φ_p, через которые можно выразить полиномы f_l, $l = 1, \ldots, L$. Параметр P при этом остается произвольным, поэтому имеет смысл придать ему последовательные значения $P = 1, 2, \ldots, P_{\min}$, пока не дойдем до минимального значения $P_{\min}$, при котором уравнение (80) имеет хотя бы одно решение. Для дальнейшего существенно, что в результате решения уравнения (80), будут найдены не только базис порождающих полиномов φ_p, но и разложение f_l по $L_i(\varphi)$, даваемые вторым равенством (72). Это существенно потому, что после определения подстановки (70) для каждого буфера разложение кодирующего полинома $F(l_i)$ по $L_j^l(e_i)$ (41) превратится в L разложений $F(e_i)$ по $L_i(\varphi)$.

Подстановки (42) или (70) могут быть заданы или в виде

$$e_i^l = \sum_{i=1}^{P} C_{ip}^l \varphi_p, \tag{81}$$

причем при фиксированном l коэффициенты C_{ip}^l удовлетворяют условию

$$\forall\, l \in \{1, \ldots, L\}\ \forall\, i \in \{i, \ldots, I\}\ \exists!\, p \in \{i, \ldots, P\}\ C_{ip}^l = 1,$$

т. е. матрица C_{ip}^l в каждой строке содержит одну и только одну единицу, а остальные — нули, или в виде

$$e_i^l = \sum_{p=0}^{2^P-1} b_{ip}^l L_p(\varphi). \tag{82}$$

Но, так как согласно (42) $e_i^l = \varphi_{l_i}$, то коэффициенты b_{ip}^l удовлетворяют условию:

$$b_{ip}^l = \begin{cases} 1, & \text{если } p = 0,\ 1,\ \ldots,\ l_i - 1,\ l_i + 1,\ \ldots,\ 2^P - 1, \\ 0, & \text{если } p = l_i, \end{cases} \tag{83}$$

т. е. матрица b_{ip}^l в каждой строке содержит один и только один нуль, а остальные — единицы. Из (83) следует, что

$$\sum_{p=0}^{2^P-1} b_{ip}^l = 1. \tag{84}$$

Из (41) будем иметь

$$F(e_i^l) = \sum_{j=0}^{2^I-1} C_j^e L_j^e(e_i^l) = \sum_{j=0}^{2^I-1} C_j^e \prod_{k=1}^{I} L_{jk}^e(e_k^l), \tag{85}$$

где j_k определяется из равенства

$$j = \sum_{k=0}^{I} j_k 2^{k-1}. \tag{86}$$

Далее,

$$\prod_{k=1}^{I} L_{jk}^l(e_k^l) = \prod_{k=1}^{I} (1 + j_k + e_k^l) = \prod_{k=1}^{I} \left[1 + j_k + \sum_{p=0}^{2^P-1} b_{ip}^l L_p(\varphi)\right] =$$

$$= \prod_{k=1}^{I} \sum_{p=0}^{2^P-1} (1 + j_k + b_{ip}^l) L_p(\varphi) = \sum_{p=0}^{2^P-1} \left[\prod_{k=1}^{I} (1 + j_k + b_{ip}^l)\right] L_p(\varphi). \tag{87}$$

Подставляя (87) в (85) и меняя порядок суммирования по j и по p, получаем:

$$F(e_i^l) = \sum_{p=0}^{2^P-1} \left\{ \sum_{j=0}^{2^I-1} \left[C_j^e \prod_{k=1}^{I} (1 + j_k + b_{ip}^l) \right] \right\} L_p(\varphi). \tag{88}$$

Подставляя (88) в (43), причем правую часть (43) записывая в виде разложения по полиномам Лагранжа согласно (72), и приравнивая коэффициенты при соответсвующих $L_p(\varphi)$, получим уравнение

$$\sum_{j=0}^{2^I-1}\left[C_j^e\prod_{k=1}^{I}(1+j_k+b_{ip}^l)\right]+a_p^l=0. \tag{89}$$

В системе уравнений (89) неизвестными булевыми переменными являются C_j^e, определяющие кодирующий полином F, и b_{ip}^l, определяющие подстановку (70). Взяв дизъюнкцию по индексам $p=0,1,\ldots,2^P-1$ и $l=1,2\ldots,L$ от левых частей (89), можно записать систему (89) в виде одного уравнения:

$$\bigvee_{p=0}^{2^P-1}\bigvee_{l=1}^{L}\left\{\sum_{j=0}^{2^I-1}\left[C_j^e\prod_{k=1}^{I}(1+j_k+b_{ip}^l)\right]+a_p^l\right\}=0. \tag{90}$$

В уравнении (90) свободным параметром является I, который по известным из разбиения файла n и L и ранее найденным $P_{\min}$ определяется либо из неравенства (46), в котором нужно вместо P взять $P_{\min}$, либо из неравенств (64), (66), (69), причем определив из двух этих неравенств интервал $[I_{\min},I_{\max}]$ следует I последовательно придавать значения $I_{\min}$, $I_{\min}+1$, $I_{\min}+2$, …, $I_{\max}$. Если уравнение (90) ни для одного из этих значений решений не имеет, то следует последовательно увеличивать $P_{\min}$, переходя к $P_{\min}+1$, $P_{\min}+2$, …, $P_{\min}+S$ до тех S, которые удовлетворяют неравенству (46).

6. Алгоритм построения поля принадлежности

Соберем вместе все этапы построения кода поля принадлежности.

1. Выбрать длину кортежа n.

2. Файл длиной N_Φ разбить на кортежи длиной n.

3. Объединить кортежи в L буферов, содержащих по m_l ($l=1,\ldots,L$) кортежей.

4. В каждом буфере оставить только по одному разному кортежу, т. ч. длины кортежей станут равными $S_l\le m_l$ ($l=1,\ldots,L$).

5. Рассматривая каждый кортеж как двоичную запись натурального числа a_i^l, $i=1,2,\ldots,s_l$, построить булевы полиномы поля принадлежности f_l по правилу:

$$f_l=\sum_{j=0}^{a_1^l-1}L_j+\sum_{j=a_1^l+1}^{a_2^l-1}L_j+\cdots+\sum_{j=a_s^l+1}^{2^n-1}L_j=\sum_{j=0}^{2^n-1}C_j^lL_j. \tag{91}$$

6. Используя C_j^l из (91), записать уравнение (80) для $P=1,\ldots,P_{\min}$, придавая параметру P последовательно указанные выше значения до тех пор, пока у уравнения (80) не появятся решения, чем и определяется $P_{\min}$.

7. Используя значения $P_{\min}$ и неравенство (46) или его следствия (64), (66), (69), определить интервал $[I_{\min}, I_{\max}]$ допустимых значений для I.

8. Записать уравнение (90) для последовательных значений

$$I = I_{\min},\ I = I_{\min} + 1,\ \ldots,\ I = I_{\max}$$

и, последовательно придавая булевым неизвестным C_j^e и b_{ij}^l значения 0 и 1, с учетом ограничения (83), найти C_j^e и b_{ij}^l.

9. Если уравнение (90) при условиях п. 8 решений не имеет, то последовательно увеличивая P, начиная со значений $P = P_{\min},\ P = P_{\min} + 1, \ldots$, повторить для этих значений P действия из пп. 6–8 до тех пор, пока у уравнения (90) не появятся решения или не нарушится неравенство (46). В первом случае для данного разбиения файла сжатие возможно. Тогда будет определено число P порождающих $\varphi_p,\ p = 1, \ldots, P$. Во втором случае надо менять разбиение файла.

10. Составить список φ_p, упорядочив их любым способом и присвоить каждому φ_p номер $m_p,\ p = 1, \ldots, P$.

11. Послать коэффициенты C_j^e и коэффициенты C_{pj} в соответсвующие ячейки шаблона общего поля.

12. В соответствующие ячейки шаблона поля принадлежности l-го буфера послать номера m_i^l порождающих полиномов φ_p из п. 10 в соответствии с правилом: если для фиксированных l и i единственный равный нулю коэффициент b_{ip}^l есть $b_{ip_i}^l$, то $m_i^l = m_{p_i}^l$.

7. Алгоритм декодирования поля принадлежности

Задача декодирования заключается в том, чтобы по коэффициентам кодирующего полинома $F(l_k)\, C_k^e$ $(k = 0,\ 1,\ \ldots,\ 2^I - 1)$, коэффициентам полиномов φ_p порождающего базиса C_{pj} $(p = 1,\ 2,\ \ldots,\ P,\ j = 0,\ 1,\ \ldots,\ 2^n - 1)$ из шаблона общего поля и номерам m_i^l $(l = 1,\ 2,\ \ldots,\ L;\ i = 1,\ 2,\ \ldots,\ I)$ базисных полиномов φ_p из кода поля принадлежности l-го буфера, которые надо поставить в соответсвующие места кортежа e_i^l, восстановить s_l кортежей, входящих в l-й буфер.

Процедура декодирования включает в себя следующие этапы.

1. По кортежу $(m_1^l,\ m_2^l,\ \ldots,\ m_I^l)$ из поля принадлежности l-го буфера, используя коэффициенты C_{pj} из шаблона общего поля, получаем согласно (44) кортеж из подмножества базиса порождающих полиномов $\left(\varphi_{m_1^l},\ \varphi_{m_2^l},\ \ldots,\ \varphi_{m_I^l}\right)$.

2. Заменяем кортеж $(e_1^l,\ e_2^l,\ \ldots,\ e_I^l)$ аргументов кодирующего полинома $F(l_i)$ $(i = 1,\ \ldots,\ I)$ на кортеж $\left(\varphi_{m_1^l},\ \varphi_{m_2^l},\ \ldots,\ \varphi_{m_I^l}\right)$.

3. Вычисляем $L_j^e(e_i)$ с заменой из п. 2 согласно условиям

$$L_j^e(e_i^l) = \prod_{k=1}^{2^I-1} L_{j_k}(e_k^l) = \prod_{k=1}^{I}(1 + j_k + e_k^l) = \prod_{k=1}^{I}(1 + j_k + \varphi_{m_k^l}), \qquad (92)$$

причем в (92) j_k определяется согласно равенству

$$j = \sum_{k=1}^{I} j_k 2^{k-1}. \tag{93}$$

4. Используя c_j^e из шаблона общего поля и (92) получаем, в силу (41), (42) и (43), что

$$F(e_k^l) = \sum_{j=1}^{2^I-1} C_j^e L_j^e(e_k^l) = \sum_{j=1}^{2^I-1} C_j^e \prod_{k=1}^{I} (1 + j_k + \varphi_{m_k^l}) = f_l(x). \tag{94}$$

5. Булево уравнение $f_l(x_i) = 0$ для l-го буфера решаем последоватлельно придавая булевым переменным x_1, x_2, ..., x_n значения 0 и 1 и, тем самым восстанавливая кортежи входящие в l-й буфер.

6. Действия в пп. 1–5 нужно выполнить последовательно придавая индексу l значения 1, 2, ..., L.

8. Алгоритм, адаптирующий код поля принадлежности к файлу

При заданном разбиении файла, определяемом числами n, L, m_l, s_l ($l = 1, 2, \ldots, L$), известны коэффициенты C_j^l в разложении

$$f_l(x) = \sum_{j=0}^{2^n-1} C_j^l L_j(x). \tag{95}$$

Если существует базис порождающих полиномов $\{\varphi_1, \varphi_1, \ldots, \varphi_P\}$, то это значит, что f_l могут быть выражены через φ_p, $p = 1, \ldots, P$, т. е.

$$f_l = f_l(\varphi_p). \tag{96}$$

Но тогда, обозначая через $L_k^{\varphi}(\varphi_p)$ ($k = 0, 1, \ldots, 2^P - 1$) полиномы Лагранжа, построенные из φ_p, можно записать (96) в виде

$$f_l(\varphi) = \sum_{j=0}^{2^P-1} C_k^l L_k^{\varphi}(\varphi_p). \tag{97}$$

Поскольку $f_l(x)$ известно из (95) и однозначно определяется файлом и его разбиением, учитывая, что $L >> P$, т. е. полиномов $f_l(x)$ намного больше, чем полиномов φ_p, то возникает задача о том, чтобы в качестве порождающего базиса взять P полиномов из известных нам L полиномов f_l, т. е. задача о выделении из множества $\{f_1, f_2, \ldots, f_L\}$ подмножества $\{f_{l_1}, f_{l_2}, \ldots, f_{l_P}\}$, которое играло бы роль порождающих базисных полиномов. Такое построение базиса порождающих полиномов логично назвать

адаптацией кода поля принадлежности к файлу. Для того, чтобы такая адаптация была возможной, нужно исследовать возможность решения системы (97) относительно φ_p ($p = 1, 2, \ldots, P$). Для этого заметим, что

$$\varphi_p = \sum_{k=0}^{p-1} L_k^{\varphi}(\varphi) + \sum_{k=p+1}^{2^P-1} L_k^{\varphi}(\varphi) = L_p^{\varphi}(\varphi) + 1. \tag{98}$$

Поскольку в правой части (97) и (98) стоят линейные комбинации $L_k^{\varphi}(\varphi)$, то очевидно, что если φ_p можно выразить через f_l, то это будет линейная функция f_l. Обозначая коэффициенты этой функции для полинома φ_p через α_l^p ($p = 1, \ldots, P$; $l = 1, \ldots, L$), умножая (97) на α_l^p и суммируя по l, получим

$$\begin{aligned}
&\sum_{l=1}^{L} \alpha_l^p f_l(\varphi) = \sum_{l=1}^{L} \alpha_l^p \sum_{k=0}^{2^P-1} C_k^l L_k^{\varphi}(\varphi) = \\
&= \sum_{k=0}^{2^P-1} \left(\sum_{l=1}^{L} \alpha_l^p C_k^l \right) L_k^{\varphi}(\varphi) = \sum_{k=0}^{p-1} \left(\sum_{l=1}^{L} \alpha_l^p C_k^l \right) L_k^{\varphi}(\varphi) + \\
&+ \sum_{l=1}^{L} \alpha_l^p C_k^l L_k^{\varphi}(\varphi) + \sum_{k=p+1}^{2^P-1} \left(\sum_{l=1}^{L} \alpha_l^p C_k^l \right) L_k^{\varphi}(\varphi) = \varphi_p,
\end{aligned} \tag{99}$$

если потребовать, чтобы (99) равнялось φ_p. Приравнивая коэффициенты при $L_k^{\varphi}(\varphi)$ в (98) и (99), получаем систему булевых уравнений для определения булевых переменных α_l^p (их PL) и коэффициентов C_k^l в разложении (97) (их $L \cdot 2^P$):

$$\sum_{l=1}^{L} \alpha_l^p C_k^l = \begin{cases} 1, & \text{если } k = 0, 1, \ldots, p-1, \\ 0, & \text{если } k \geq p. \end{cases} \tag{100}$$

Система (100) содержит $P \cdot 2^P$ булевых уравнений. Их можно превратить в одно, если взять дизъюнкцию по всем значениям индексов k, p:

$$\bigvee_{p=1}^{P} \left\{ \bigvee_{k=1}^{p-1} \left(\sum_{l=1}^{L} \alpha_l^p C_k^l + 1 \right) \bigvee \sum_{l=1}^{L} \alpha_l^p C_k^l \bigvee_{k=p+1}^{2^P-1} \left(\sum_{l=1}^{L} \alpha_l^p C_k^l + 1 \right) \right\} = 0. \tag{101}$$

Если ввести обозначение

$$\beta_k^p = \sum_{l=1}^{L} \alpha_l^p C_k^l, \tag{102}$$

то из (101) и (102) получим

$$\bigvee_{p=1}^{P} \left\{ \bigvee_{k=1}^{p-1} (\beta_k^p + 1) \bigvee \beta_p^p \bigvee_{k=p+1}^{2^P-1} (\beta_k^p + 1) \right\} = 0. \tag{103}$$

Так как для булевых переменных, в т. ч. и для булевых полиномов, справедливы формулы

$$(f_k+1)\bigvee_{k=1}^{P}(g_k+1)=\prod_{k=1}^{P}f_kg_k+1, \tag{104}$$

$$(f_k+1)\bigvee_{k=1}^{P}g_k=f_k\bigvee_{k=1}^{P}g_k+1, \tag{105}$$

то, используя (104) и (105), приведем (103) к виду:

$$\prod_{p=1}^{P}\left\{\prod_{k=0}^{2^P-1}\beta_k^p+\prod_{k=0}^{p-1}\beta_k^p\prod_{m=p+1}^{2^P-1}\beta_m^p+\beta_p^p\right\}+1=0. \tag{106}$$

Если ввести обозначение

$$A_p=\prod_{k=1}^{p-1}\beta_k^p\prod_{k=p+1}^{2^P-1}\beta_k^p, \tag{107}$$

то (106) перепишется так:

$$\prod_{p=1}^{P}\left\{A_p\beta_p^p+A_p+\beta_p^p\right\}=1. \tag{108}$$

Но (108) будет выполняться только тогда, когда каждый множитель в (108) равен 1, т. е.

$$A_p\beta_p^p+A_p+\beta_p^p=A_p\bigvee\beta_p^p=1,\quad p=1,\ldots,P. \tag{109}$$

Уравнение (109) имеет три решения:

$$A_p=0,\ \beta_p^p=1, \tag{a}$$

$$A_p=1,\ \beta_p^p=0, \tag{b}$$

$$A_p=1,\ \beta_p^p=1. \tag{c}$$

В случае решения (c) вследствие (107) будем иметь

$$\forall\, p\in\{1,\ldots,P\}\ \forall\, k\in\{1,\ldots,2^P-1\}\ :\ \beta_k^p=1,$$

а значит из (102) получим следующую систему булевых уравнений:

$$\sum_{l=1}^{L}\alpha_l^pC_k^l+1=0,\ p=1,\ldots,P;\ k=0,\ldots,2^P-1. \tag{110}$$

В случае решения (b) вследствие (107) будем иметь

$$\forall\, p \in \{1, \ldots, P\}\ \forall\, k \in \{1, \ldots, p-1, p+1, \ldots, 2^P - 1\} : \ \beta_k^p = 1;\ \beta_p^p = 0, \tag{111}$$

а значит (102) дает:

$$\begin{aligned} &\sum_{l=1}^{L} \alpha_l^p C_p^l = 0, && p = 1, \ldots, P, \\ &\sum_{l=1}^{L} \alpha_l^p C_k^l + 1 = 0, && p = 1, \ldots, P, \\ & && k = 0, \ldots, p-1, p+1, \ldots, 2^p - 1. \end{aligned} \tag{112}$$

В случае решения (a) уравнение $A_p = 0$ с A_p из (107) будет иметь 2^{2^P-1} различных решений. Любые P решений системы (111) или (112) или аналогичных систем для любого решения 1 из (107), в каждом из которых хотя бы одно $\alpha_k^\beta = 1$, дадут согласно (99) базис порождающих полиномов φ_p, линейно выраженный через известные полиномы $f_l(x)$, а значит дадут $\varphi_p = \varphi_p(x)$. Так как $L >> P$, то выбирая P полиномов f_l из L в качестве порождающих полиномов, для остальных $L - P$ полиномов f_l получим явные, вообще говоря, нелинейные зависимости, подставив C_k^e, найденные из систем (11) или (112) и φ_p из (99) в (97).

Таким образом, задача об адаптации кода поля принадлежности к файлу в принципе может быть решена.

Список использованной литературы

1. *Ватолин Д., Ратушняк А., Смирнов М., Юдин В.* Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: Диалог - МИФИ, 2002. – 384 с.
2. *Владимиров Д. А.* Булевы алгебры. – М.: Наука, 1969. – 320 с.
3. *Гоппа В. Д.* Введение в алгебраическую теорию информации. – М.: Наука, 1995. – 112 с.
4. *Жегалкин И. И.* О технике вычислений предположений в символической логике // Матем. сб. – 1927. – № 34. – с. 9–28.
5. *Кричевский Р. Е.* Сжатие и поиск информации. – М.: Радио и связь, 1989. – 168 с.
6. *Моисил Дж.* Алгебраическая теория дискретных автоматических устройств. – М.: ИЛ, 1963. – 415 с.
7. *Толстопятов А. А.* О структуре дискретной информации и общих условиях ее сжатия // Вестник ИвГУ. – 2002. – Вып. 3. – С. 80–82.
8. *Толстопятов А. А.* О возможности использования булевых уравнений для сжатия файлов // Вестник ИвГУ. – 2003. – Вып. 3. – С. 82–84.
9. *Толстопятов А. А.* Вычисление длины поля кратности при булевом сжатии файлов // Вестник ИвГУ. – 2004. – Вып. 3. – С. 71–76.
10. *Толстопятов А. А.* Быстрый алгоритм кодирования и декодирования поля порядка при булевом сжатии файлов // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2007. – Вып. 1 (4). – С. 35–46.

11. *Толстопятов А. А.* Медленный алгоритм кодирования и декодирования поля кратности при булевом сжатии файлов // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2007. – Вып. 1 (4). – С. 47–52.

12. *Толстопятов А. А.* Быстрый алгоритм кодирования и декодирования поля кратности при булевом сжатии файлов // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2007. – Вып. 1 (4). – С. 53–78.

13. *Толстопятов А. А.* Факторы адаптации при булевом сжатии файлов // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2007. – Вып. 1 (4). – С. 79–92.

14. *Толстопятов А. А., Хашин С. И.* Алгоритм построения поля порядка при булевом сжатии // Вестник ИвГУ. – 2004. – Вып. 3. – С. 139–143.

15. *Хаффман Д. А.* Метод построения кодов с минимальной избыточностью. // Кибернетический сборник. – 1961. – Вып. 3. – С. 79–87.

16. *Шеннон К.* Математическая теория связи // Работы по теории информации и кибернетике. – М.: ИЛ, 1963. – С. 243–332.

17. *Яглом А. М., Яглом И. М.* Вероятность и информация. – М.: Наука, 1973. – 512 с.

18. *Lunch T. J.* Data Compression Techniques and Applications: Lifetime. Learning Publications. – Belmont, 1985. – 345 p.

Поступила в редакцию 29.12.2008.