РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
А В Т О М Е Т Р И Я
№ 5 1996

УДК 519.725

О. В. Мазуров
(Новосибирск)

АЛГЕБРАИЧЕСКОЕ ДЕКОДИРОВАНИЕ КЛАССА КОДОВ РИДА — СОЛОМОНА ПРИ ЧИСЛЕ ОШИБОК, БОЛЬШЕМ ПОЛОВИНЫ КОДОВОГО РАССТОЯНИЯ

Рассматривается код Рида — Соломона над полем $GF(p^q)$, имеющим нетривиальное подполе $GF(p^{q'})$, $q = q'm$. Выбор локаторов кода из этого подполя дает дополнительные возможности для декодирования. Так оказывается возможным декодирование (негарантированное), когда число ошибок не превышает $\frac{m}{m+1}(d-1)$.

Коды Рида — Соломона (РС), являясь кодами с максимально достижимым расстоянием, находят широкое применение в системах передачи и хранения информации. Они используются как сами по себе, так и во всевозможных конструкциях построения длинных кодов, например в каскадных и итеративных кодах. Известно, что коды РС обладают избыточностью, позволяющей проводить декодирование и при числе ошибок, большем $\left[\frac{d-1}{2}\right]$, где d — кодовое расстояние; однако реализовать эту способность в виде эффективного в вычислительном смысле алгоритма, насколько известно автору, до сих пор не удавалось. Оказывается, что для класса кодов специального вида, остающегося достаточно широким и интересным с практической точки зрения, такой алгоритм может быть предложен. Отметим, что узким местом при декодировании кодов РС (равно как и других кодов) является нахождение позиций ошибок в принятом слове. В данной статье рассматривается только этот шаг алгоритма, причем получение полного алгоритма декодирования, использующего предлагаемый метод, считается очевидным.

Зададим (n, k, d) код Рида — Соломона над конечным полем $GF(p^q)$ как совокупность кодовых слов X длины n, удовлетворяющих $d-1$ уравнениям:

$$\sum_{i=0}^{n-1} X_i Z_i^l = 0, \qquad l = 0, \ldots, d-2, \tag{1}$$

где $d = n - k + 1$ — кодовое расстояние; Z_i — заданные различные элементы поля $GF(p^q)$.

Из (1) следует, что для любого многочлена $P(Z)$ над $GF(p^q)$ степени не выше $d-2$ справедливо равенство

$$\sum_{i=0}^{n-1} X_i P(Z_i) = 0. \tag{2}$$

Пусть $X' = X + e$ — кодовое слово, содержащее ошибки; причем число компонент вектора ошибки e, отличных от нуля, не превосходит $d - 2$. Тогда для многочлена локаторов ошибок $P_e(Z)$ должно выполняться равенство

$$\sum_{i=0}^{n-1} X_i' P_e(Z_i) = \sum_{j=0}^{l} P_j \sum_{i=0}^{n-1} X_i' Z_i^j = \sum_{j=0}^{l} P_j S_j = 0,$$

где S_j — синдромы слова X'; $P_e(Z) = \sum_{j=0}^{l} P_j Z^j$, $P_l = 1$.

Очевидно, аналогичные равенства должны выполняться для многочленов $P_e(Z)Z$, $P_e(Z)Z^2$ и т. д., пока степень произведения остается меньше $d - 1$:

$$\sum_{j=0}^{l} P_j S_{j+k} = 0. \qquad k = 0, \ldots, d - 2 - l. \tag{3}$$

Эти равенства могут служить основой для переборного метода декодирования слова X'. Пробуя различные комбинации Z_i и проверяя выполнимость равенств (3), мы можем определить ближайшее к X' кодовое слово (или все, если их окажется несколько).

Любой переборный алгоритм определения позиций ошибок имеет полиномиальную относительно n вычислительную сложность. Можно предлагать различные улучшения, но даже квадратичная сложность при больших n лишает такие алгоритмы интереса с практической точки зрения. Альтернативой переборным методам являются алгебраические методы декодирования, имеющие линейную относительно n вычислительную сложность. Так, традиционный алгоритм декодирования (см., например, [1, 2]) при числе ошибок, меньшем половины кодового расстояния, имеет линейную по n вычислительную сложность в части вычисления синдромов и определения корней многочлена локаторов ошибок (процедура Чена).

Отметим пограничный случай, когда переборный по сути метод имеет линейную по n сложность (так сказать, полиномиальную степени 1). Так, при четном d мы можем, перебирая по очереди все Z_i и рассматривая их как стирания, решать задачу декодирования при наличии ошибок и одного стирания. При этом максимальное число ошибок, которое мы можем декодировать таким способом, становится равным $d/2$, т. е. на одну больше, чем при простом использовании алгоритма декодирования.

Теперь опишем метод, позволяющий для достаточно широкого и с практической точки зрения интересного класса кодов Рида — Соломона определять значения ошибок, когда их число превышает половину кодового расстояния, и имеющий при этом линейную по n вычислительную сложность.

Основная идея метода заключается в выборе значений Z_i из некоторого подполя основного поля $GF(p^q)$. При этом мы сильно ограничиваем максимальную длину такого кода, но получаем дополнительную возможность в определении многочлена локаторов ошибок. Отметим, что подобные коды рассматривались и ранее, например в системах каскадного кодирования [2], где они назывались укороченными кодами, однако возможности их более сильного декодирования не были отмечены.

Пусть $q = q'm$ и α — первообразный элемент $GF(p^q)$. Тогда $\beta = \alpha^{(p^q - 1)/(p^{q'} - 1)}$ — первообразный элемент подполя $GF(p^{q'})$, которое, таким образом, состоит из элементов $\{0, \beta, \beta^2, \beta^3, \ldots, \beta^{p^{q'} - 1} = 1\}$. Пусть теперь Z_i — различные элементы поля $GF(p^{q'})$. Коэффициенты любого многочлена, все корни которого лежат в $GF(p^{q'})$, также принадлежат этому подполю.

Возведем каждое уравнение из (3) в степень $p^{q'}$:

$$\left(\sum_{j=0}^{l} P_j S_{j+k}\right)^{p^{q'}} = \sum_{j=0}^{l} P_j^{p^{q'}} S_{j+k}^{p^{q'}} = \sum_{j=0}^{l} P_j S_{j+k}^{p^{q'}} = 0, \qquad k = 0, \ldots, d-2-l. \quad (4)$$

Поскольку синдромы в общем случае принадлежат основному полю, то $S_j^{p^{q'}} \neq S_j$, и мы имеем для определения неизвестных P_j дополнительно $d - l - 1$ уравнений, которые, опять же в общем случае, линейно независимы с уравнениями из (3).

Производя дальнейшее возведение в степень $p^{q'}$ уравнений (4), получим новую систему дополнительных уравнений относительно P_j. Всего вместе с (3) мы можем получить m систем по $d - l - 1$ уравнений в каждой. Требование того, чтобы число неизвестных не превышало число уравнений, выполняется при $l \leq m(d - 1 - l)$, или

$$l \leq \frac{m}{m+1}(d-1), \quad (5)$$

что при $m \geq 2$ заметно больше половины кодового расстояния.

Пример. При $d = 7$, $m = 2$ имеем возможность восстановить четыре ошибки, используя следующую систему:

$$\begin{pmatrix} S_0 & S_1 & S_2 & S_3 \\ S_1 & S_2 & S_3 & S_4 \\ S_0^{p^{q'}} & S_1^{p^{q'}} & S_2^{p^{q'}} & S_3^{p^{q'}} \\ S_1^{p^{q'}} & S_2^{p^{q'}} & S_3^{p^{q'}} & S_4^{p^{q'}} \end{pmatrix} \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{pmatrix} = \begin{pmatrix} S_4 \\ S_5 \\ S_4^{p^{q'}} \\ S_5^{p^{q'}} \end{pmatrix}.$$

Описанный метод является вариантом метода, предложенного в [3], и связь между ними может быть установлена следующим образом. Поскольку поле $GF(p^q)$ — расширение поля $GF(p^{q'})$, то для некоторых фиксированных $C_0, C_1, \ldots, C_{m-1} \in GF(p^q)$ любой элемент $X_i \in GF(p^q)$ представим в виде

$$X_i = Y_{i,0}C_0 + Y_{i,1}C_1 + \ldots + Y_{i,m-1}C_{m-1}, \quad (6)$$

где $Y_{i,j} \in GF(p^{q'})$, причем такое представление единственно. В качестве C_j можно, например, выбрать $\alpha^{jp^{q'}}$. В силу линейности и единственности такого представления код (1) распадается на совокупность m кодов над $GF(p^{q'})$:

$$\sum_{i=0}^{n-1} Y_{i,j} Z_i^l = 0, \qquad l = 0, \ldots, d-2, \quad j = 0, \ldots, m-1, \quad (7)$$

причем ошибка в компоненте X_i (даже одиночная битовая) приводит, как правило, к ошибкам по всем (или, по крайней мере, по многим) соответствующим компонентам $Y_{i,j}$, что и делает применимым в таком варианте метод, описанный в [3].

Несколько слов о вероятностях ошибки декодирования и отказа от декодирования предложенного метода. Вероятность ошибочного декодирования описанного метода несколько больше, чем у традиционного алгоритма, просто в силу того, что декодированию поддается значительно большее число слов X'. На вопрос, имеет ли метод практический смысл, т. е. какая доля слов X',

действительно, поддается декодированию при числе ошибок, достигающих предела (5), ответ можно дать на основании результатов вычислительного эксперимента. Так, для случая основного поля $GF(2^{16})$, подполя локаторов $GF(2^8)$, $n = 256$, $d = 4, 7, 10, 13, 16$ и числа ошибок соответственно $l = 2, 4, 6, 8, 10$ вычислительная оценка вероятности отказа от декодирования составила менее 0,5 % (число экспериментов ~100 000 для каждой пары d, l). При этом не произошло ни одного случая ошибочного декодирования.

В заключение интересно сравнить рассмотренные коды с обычными кодами над $GF(p^q)$. Такое рассмотрение, не уменьшая общности выводов, проведем для конкретных двоичных кодов. Логично для сравнения выбрать два кода, имеющих одинаковую длину и избыточность, выраженную в битах. Так, например, укороченный (128, 116, 13) код над $GF(2^{16})$ и обычный (256, 232, 25) код над $GF(2^8)$ имеют одинаковую битовую длину — 2048 бит и избыточность — 192 бит. Отметим сразу, что максимальная длина укороченного кода для рассматриваемых полей в 2 раза превышает максимальную длину простого кода. В части исправления одиночных битовых ошибок укороченный код, исправляющий с вероятностью, близкой к единице, восемь ошибок (в силу (5)), уступает простому коду, способному исправить 12 ошибок. Он, однако, выглядит лучше в части исправления пакетов ошибок. Так, укороченный код с вероятностью, близкой к единице, восстанавливает пакет длиной $8 \cdot 16 = 128$ бит, в то время как простой код ограничен максимальной длиной пакета в $8 \cdot 12 = 96$ бит. Таким образом, появляется возможность выбора способа кодирования в зависимости от характеристик канала или требований по восстановительной способности.

СПИСОК ЛИТЕРАТУРЫ

1. **Питерсон У., Уэлдон Э.** Коды, исправляющие ошибки. М.: Мир, 1976.
2. **Форни Д.** Каскадные коды. М.: Мир, 1970.
3. **Мазуров О. В.** Декодирование кода Рида — Соломона для векторов в системах хранения информации // Автометрия. 1995. № 6.

Поступила в редакцию 18 марта 1996 г.