

УДК 681.391

НЕДВОИЧНЫЕ МНОГОПороГОВЫЕ ДЕКОДЕРЫ И ДРУГИЕ МЕТОДЫ КОРРЕКЦИИ ОШИБОК В СИМВОЛЬНОЙ ИНФОРМАЦИИ ДЛЯ СИСТЕМ ПЕРЕДАЧИ И ХРАНЕНИЯ ДАННЫХ

Н.А. Кузнецов, В.В. Золотарёв, Г.В. Овечкин, П.В. Овечкин

Аннотация: Анализируются возможности основных современных недвоичных кодов и алгоритмов их декодирования, таких как коды Рида-Соломона, недвоичные низкоплотностные коды и недвоичные самоортогональные коды. Показано, что в настоящее время наиболее эффективными являются подходы к коррекции символьных ошибок, основанные на применении недвоичных многопороговых декодеров (q МПД) самоортогональных кодов. Представлены характеристики q МПД для новых кодовых схем с параллельным кодированием, обеспечивающих эффективную работу при существенно большем уровне шума в канале. Описано применение q МПД для защиты файлов от искажений.

Ключевые слова: помехоустойчивое кодирование, системы передачи данных, системы хранения данных, q -ичный симметричный канал, недвоичные коды, коды Рида-Соломона, недвоичные самоортогональные коды, недвоичный многопороговый декодер, недвоичные низкоплотностные коды

Введение

Помехоустойчивое кодирование применяется для исправления ошибок, возникающих при передаче данных по каналам с шумами или при их длительном хранении на различного рода носителях информации. В настоящее время в литературе наибольшее внимание уделяется методам коррекции ошибок в двоичных данных. Однако во многих реальных системах проще обрабатывать данные, имеющие байтовую структуру. Например, удобнее работать с байтами в системах хранения больших объемов информации (оптические диски и др. носители). В подобных системах для защиты данных от ошибок целесообразно применение недвоичных помехоустойчивых кодов. Кроме этого, недвоичные коды оказываются эффективнее лучших двоичных в каналах с пакетирующимися ошибками, поскольку при даже большом пакете ошибок в двоичных данных искаженными час-

то оказываются всего несколько недвоичных символов, легко исправляемых декодером недвоичного кода.

На сегодняшний день в теории кодирования известен ряд недвоичных кодов, различающихся корректирующей способностью, вносимой избыточностью, сложностью декодирования и многими другими важными параметрами. Целью данного обзора является изложение возможностей основных, наиболее эффективных недвоичных кодов и методов их декодирования.

Рассмотрим эффективность современных методов коррекции ошибок в символьных данных при различных параметрах кода и размерах символа. При сравнении характеристик будем использовать классическую модель q -ичного симметричного канала (q СК), которая хорошо подходит для оценивания возможностей данных методов. В таком канале каждый символ искажается независимо с вероятностью P_0 , причем при искажении символ с равной вероятностью переходит в один из $q-1$ других символов. Подобная модель, например, соответствует каналу с пакетами ошибок при использовании перемежения/деперемежения на уровне символов.

Одной из основных характеристик канала связи является его пропускная способность C , которая определяет максимальную кодую скорость используемого кода, при которой с помощью правильно выбранных кодера и декодера можно вести передачу по каналу с шумом со сколь угодно малой вероятностью ошибки. Для q СК пропускная способность канала определяется как [1]

$$C_{qCK} = 1 + P_0 \cdot \log_q(P_0) + (1 - P_0) \cdot \log_q(1 - P_0) - P_0 \cdot \log_q(q - 1), \quad (1)$$

где P_0 – вероятность ошибки в канале.

Из данного выражения также можно определить максимальную вероятность ошибки, при которой способен работать код с заданной кодовой скоростью.

1. Коды Рида-Соломона

На сегодняшний день среди двоичных кодов практическое применение нашли только коды Рида-Соломона (РС) [2], обладающие рядом положительных свойств. Коды РС характеризуются тем, что для исправления в пределах кодового слова любой комбинации из t символьных ошибок можно использовать лишь $2t$ проверочных символов. При этом длина кода n у обычных кодов РС должна быть строго меньше размера алфавита q . Для коротких кодов РС существуют эффективные алгоритмы декодирования, в полной мере использующие корректирующие возможности кода [3, 4]. Сложность реализации наиболее простых из них пропорциональна $n \cdot \log^2 n$ [4]. Под сложностью реализации здесь и далее понимается число арифметических операций, требуемых для декодирования кодового блока.

Характеристики кодов РС с кодовой скоростью $R=1/2$ и длиной $n=255$ однобайтовых символов (размер алфавита $q=256$) в q СК представлены на рис. 1 кривой 1. По оси абсцисс на рисунке отложена вероятность ошибки P_0 в q СК, а по оси ординат – оценка вероятности ошибки на символ после декодирования, полученная путем компьютерного моделирования. При этом из (1) можно получить, что теоретически в q СК для $q=256$ и $R=1/2$ можно работать при $P_0=0.380$. Видно, что показываемые кодами РС характеристики очень далеки от теоретически возможных. Потенциальные возможности более длинных кодов РС с кодовой скоростью $R=1/2$ и длиной $n=65535$ двухбайтовых символов ($q=2^{16}$) отражены на рис. 1 кривой 2. И такие коды работают при вероятности ошибки в канале, значительно меньшей теоретически возможной, равной $P_0=0.438$ для данных условий.

Особый интерес для систем передачи и хранения данных часто представляют малоизбыточные помехоустойчивые коды. Для таких кодов РС характеристики представлены на рис. 2. Здесь кривая 1 отражает вероятность символьной ошибки декодера кода РС с кодовой скоростью $R=7/8$ для однобайтовых символов ($q=256$). Отметим, что теоретически при данных R и q можно работать при вероятности ошибки в канале около 0.076. Характеристики кодов РС с еще большей кодовой

скоростью $R=19/20$ при $q=256$ представлены на рис. 2 кривой 2. Для указанных параметров канала и кодирования предельно возможный уровень шума, при котором теоретически возможно эффективное декодирование, составляет $P_0 \approx 0.027$.

Заметим, что кроме кодов РС в настоящее время вообще нет других коротких недвоичных кодов, имеющих достаточно эффективные и одновременно простые методы декодирования. Однако короткие коды РС длины до $n=255$ однобайтовых символов, как следует из рис. 1 и 2, не обеспечивают необходимых в настоящее время уровней достоверности. А декодеры для длинных кодов РС оказываются слишком сложными для реализации, так как их сложность пропорциональна $n \cdot \log^2 n$ и возможное существенное упрощение данных декодеров весьма проблематично. Отметим, что для кодов РС существуют алгоритмы декодирования, позволяющие исправлять даже несколько большее, чем t , число ошибок, например, алгоритм Судана [5]. Данные методы очень интересны для теории кодирования, однако сложность их реализации становится пропорциональной n^3 , в то время как рост эффективности декодирования от такого усложнения, особенно при высоких кодовых скоростях, которые часто и требуется применять на практике, оказывается незначительным [6]. Это иллюстрируется кривой 3 на рис. 1, который соответствует алгоритму декодирования Судана для кода РС с $n=255$, $q=256$ и $R=1/2$.

2. Недвоичные многопороговые декодеры

Гораздо ближе к теоретическим границам работают q -ичные многопороговые декодеры (q МПД) [7, 8, 9, 10] недвоичных самоортогональных кодов. Они, как и обычные двоичные МПД [8], обладают свойством стремления к решению оптимального декодера при линейной от длины кода сложности реализации, которая свойственна только пороговым процедурам. В отличие от кодов РС для q МПД никаких ограничений по длине кода вообще нет, поскольку длина кода n и размер алфавита q в недвоичных кодах с мажоритарным декодированием совершенно не

зависят друг от друга. При этом сложность декодирования кодового блока пропорциональна $n \cdot d \cdot I$, где n – длина кода, d – кодовое расстояние (обычно $d \leq 20$), I – число итераций декодирования (обычно $I \leq 30$).

Характеристики q МПД также представлены на рис. 1. Здесь кривыми 4 и 5 показана эффективность q МПД для самоортогональных кодов с $R=1/2$ и длиной блока 4000 и 32000 однобайтовых символов ($q=256$). Объем моделирования в нижних точках данных графиков составлял от $5 \cdot 10^{10}$ до $2 \cdot 10^{12}$ символов, что свидетельствует о крайней простоте метода. Из рисунка видно, что характеристики q МПД оказываются гораздо лучше характеристик кодов РС с такими же q и R . При увеличении длины блока, что для q МПД не вызывает никаких сложностей, разница в эффективности становится еще более существенной. Характеристики q МПД при использовании двухбайтовых символов представлены на рис. 1 кривой 6. Здесь также использовался код с $R=1/2$ и $n=32000$ символов. Отметим, что очень простой для реализации q МПД для двухбайтового кода длины 32000 оказывается способным обеспечить помехоустойчивость, недостижимую даже для кода РС длины 65535 двухбайтовых символов, декодер для которого на данный момент слишком сложен для реализации. При этом q МПД для двухбайтовых символов практически ни в чем не сложнее однобайтового, так как его сложность совершенно не зависит от размера алфавита q .

Высокой корректирующей способностью обладают и q МПД для малоизбыточных недвоичных самоортогональных кодов, пример характеристик которых для $R=7/8$, $n=48000$ символов и $q=256$ представлен на рис. 2 кривой 3. Здесь также видно заметное преимущество q МПД над кодами РС. Аналогичная ситуация наблюдается и при использовании кодов с еще более высокой кодовой скоростью $R=19/20$. Для данной кодовой скорости при $q=256$ эффективность q МПД показана кривой 4, а для кодов РС – кривой 2. Такие же высокие характеристики обеспечивает q МПД малоизбыточных кодов при использовании алфавита большего объема, при котором создание других декодеров представляется очень сложным. На

рис. 2 кривой 5 представлена эффективность q МПД для кода с $R=7/8$ при использовании двухбайтовых символов ($q=65536$).

Отметим, что для достижения с помощью q МПД таких результатов требуется очень тщательно выбирать применяемые коды, основным критерием при отборе которых является степень устойчивости к эффекту размножения ошибок [8], который проявляется в том, что после первой ошибки декодирования существенно увеличивается вероятность последующих ошибок. Известно, что размножению ошибок в наименьшей степени подвержены коды для схем с параллельным кодированием [11]. В [12] показано, что оптимизируя структуру данных кодов можно еще улучшить эффективность работы q МПД. В частности, характеристики найденных в [12] кодов с $q=256$ и кодовыми скоростями $1/2$ и $7/8$ представлены на рис. 1 и 2 кривыми 7 и 6 соответственно. Видно, что данные коды обеспечивают эффективную работу при больших вероятностях ошибки в q СК, чем ранее представленные [9], при такой же сложности их декодирования.

Согласно общим принципам теории кодирования, переход к каскадным принципам кодирования еще более улучшит характеристики q МПД без существенного усложнения декодера. В [9] показано, что применение совместно с q МПД простейшего кода с контролем по модулю q позволяет на несколько порядков снизить вероятность ошибки на блок по сравнению с обычным q МПД при всего лишь 2% росте избыточности [6, 8]. При этом увеличение объема вычислений в каскадном коде составляет менее 20% по сравнению с исходным алгоритмом q МПД.

Таким образом, недвоичный аналог алгоритма МПД может обеспечить при весьма высоких уровнях шума вероятности ошибки декодирования, в ряде случаев недоступные для кодов Рида-Соломона сколько угодно большой длины. При этом сложность реализации такого алгоритма оказывается незначительной, линейно растущей с длиной кода, т.е. теоретически минимально возможной [9].

3. Недвоичные низкоплотностные коды

В последнее время зарубежными специалистами стали активно развиваться декодеры недвоичных низкоплотностных (q LDPC) кодов [13]. Данные методы, безусловно, обладают очень высокой корректирующей способностью, однако сложность их реализации при больших значениях основания кода q оказывается слишком большой для практического применения в реальных системах. В частности сложность одной итерации декодирования кодового блока для одного из наиболее простых из известных алгоритмов декодирования q LDPC кодов пропорциональна $n \cdot q \cdot \log_2 q$ [14]. В результате разница в сложности реализации q МПД и декодеров q LDPC кодов при использовании всего четырехбайтовых символов ($q=2^{32}$) превышает миллиард раз. q МПД же при этом будет иметь ту же символьную скорость работы, как и для однобайтовых символов, а его битовая производительность даже возрастет в 4 раза. В [15] предложен метод декодирования q LDPC кодов, который обладает сложностью, пропорциональной $n \cdot s^2$, где $s \leq q$ – максимальный размер списка, передаваемого по ветвям графа q LDPC кода и содержащего наиболее вероятные символы, соответствующие этим ветвям. Такое ограничение размера списка существенно упрощает процесс декодирования, но приводит к некоторому ухудшению характеристик. Например, для размера алфавита $q=2^{32}$ декодер регулярного q LDPC кода длиной 100000 символов и кодовой скоростью $R=1/2$ при $s=q$ теоретически способен работать при вероятности ошибки в канале $P_0=0.429$, а при $s=32$ работает только при $P_0=0.232$ (пунктир 8 на рис. 1) [15]. Следует особо отметить что декодер q LDPC кода с $s=32$ обладает существенно меньшей корректирующей способностью, чем q МПД при символах такого же размера (кривая 9 на рис. 1), и примерно в тысячу раз большей вычислительной сложностью.

4. Применение q МПД в системах хранения данных

Одной из областей применения недвоичных кодов является защита данных от искажений при долговременном хранении на различных носителях информации. Для этих целей в настоящее время используются такие программные пакеты, как QuickPar [16] и ICE ECC [17], основанные на применении кодов РС. При работе данных пакетов с большими файлами возникают сложности или с обеспечением приемлемой скорости, или надежности исправления ошибок. Применение для защиты файлов программных средств [18], использующих алгоритмы q МПД, решает перечисленные проблемы, часто предоставляя одновременно и большую корректирующую способность, и гораздо более высокое быстродействие. В частности, q МПД при программной реализации даже для длинных кодов и больших размеров алфавита обеспечивают скорость декодирования в несколько десятков Мбит/с на обычном ПК [10], что оказывается в десятки, сотни, а иногда и в тысячи раз быстрее других современных алгоритмов коррекции ошибок. Такие скорости, например, показывает представленная на сайте www.mtdbest.iki.rssi.ru демопрограмма для q МПД, работающая даже на обычных ПК на скоростях 8...30 Мбит/с при столь больших шумах канала, при которых декодеры кодов РС вообще не работают хоть сколь-нибудь эффективно. В результате использующие q МПД программные средства в ряде случаев способны обеспечить на много порядков более высокие уровни защиты файлов от искажений, чем указанные выше программы, поддерживая при этом во много раз лучшие скорости кодирования и восстановления информации [18].

Сравнение возможностей программ для защиты файлов от искажений показало, что применение q МПД позволяет существенно повысить скорость кодирования/восстановления информации по сравнению с аналогами. Программные средства, основанные на q МПД, обеспечивали скорость кодирования/декодирования в десятки мегабайт в секунду, что в десятки раз больше скорости работы программ ICE ECC и QuickPar в тех же условиях. Особо отметим, что при этом q МПД одинаково эффек-

тивно исправляет как независимые ошибки и стирания, так и пакеты ошибок или стираний. Этого нельзя сказать о программах ICE ECC, QuickPar, которые эффективно исправляют пакеты ошибок, но не справляются даже с малым процентом независимых ошибок.

Заключение

Для исправления ошибок в символьных данных в настоящее время практически везде используются коды Рида-Соломона, хотя их эффективность достаточно далека от теоретически возможной. Конкуренцию данным кодам должны составить недвоичные самоортогональные коды, недвоичные многопороговые декодеры для которых оказываются на много порядков лучше декодеров кодов Рида-Соломона как по вероятности ошибки, так и по числу операций декодирования. Уникальность характеристик недвоичных МПД позволяет говорить об их абсолютном преимуществе над декодерами кодов Рида-Соломона для всех достаточно длинных кодов при любых параметрах кодирования. Это определяется эффективным переносом идей многопорогового декодирования на очень просто организованные недвоичные самоортогональные коды сколь угодно большой длины. Практическое применение недвоичных низкоплотностных кодов при больших размерах символа будет проблематичным из-за значительной сложности реализации декодеров данных кодов.

На веб-сайте www.mtdbest.iki.rssi.ru представлены демопрограммы, иллюстрирующие работу основных методов коррекции ошибок. Используя данные демопрограммы можно оценить эффективность и скорость алгоритмов декодирования недвоичных помехоустойчивых кодов.

Работа выполнена при финансовой поддержке РФФИ (грант №08-07-00078).

Список литературы

1. Low-Density Parity-Check Codes with Rates Very Close to the Capacity of the q -ary Symmetric Channel for Large q // ISIT 2004, Chicago, USA, June 27 – July 2, 2004. pp. 273.
2. Reed I.S., Solomon G. Polynomial codes over certain finite fields // J. Soc. Industrial Appl. Math., 1960, vol.8, pp.300–304.
3. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки. М.: Мир, 1976.
4. Ning C., Zhiyuan Y. Complexity analysis of Reed-Solomon decoding over $GF(2^m)$ without using syndromes // EURASIP Journal on Wireless Communications and Networking, January 2008, n.4, pp.1-11,
5. Sudan M. Decoding of Reed Solomon codes beyond the error-correction bound // Journal of Complexity, 1997, vol.13, pp.180–193.
6. Золотарёв В.В. Каскадные схемы МПД-декодирования для больших баз данных. Мобильные системы, 2008, №3, С.66-71.
7. Золотарёв В.В., Овечкин Г.В. Эффективное многопороговое декодирование недвоичных кодов. Радиотехника и электроника. В печати.
8. Золотарёв В.В. Теория и алгоритмы многопорогового декодирования. М.: Радио и связь, Горячая линия – Телеком, 2006.
9. Золотарёв В.В. Обобщение алгоритма МПД на недвоичные коды. Мобильные системы, 2007, №3, С.39–42.
10. Многопороговые декодеры. Веб-сайт ИКИ РАН www.mtdbest.iki.rssi.ru.
11. Золотарёв В.В. Параллельное кодирование в каналах СПД // Вопросы кибернетики. – 1986. – Вып. 120.
12. Овечкин Г.В., Овечкин П.В. Оптимизация структуры недвоичных самоортогональных кодов для схем параллельного кодирования // Труды НИИР, 2009. №2.
13. Davey M.C., MacKay D.J.C. Low density parity check codes over $GF(q)$ // IEEE Comm. Letters, 2(6), 1998, pp.165–167.

14. Declercq D., Fossorier M. Extended minsum algorithm for decoding LDPC codes over $GF(q)$ // IEEE International Symp. on Inf. Theory, 2005, pp.464–468.

15. Zhang F., Pfister H. List-Message Passing Achieves Capacity on the q -ary Symmetric Channel for Large q // In Proc. IEEE Global Telecom. Conf., Washington, DC, Nov. 2007. pp.283–287.

16. www.quickpar.org.uk

17. www.ice-graphics.com

18. Овечкин П.В. Применение недвоичного многопорогового декодера для защиты файлов от искажений // В сб.: «11 Международная конференция «Цифровая обработка сигналов и ее приложения- DSPA-09», М., 2009. С.200–202.

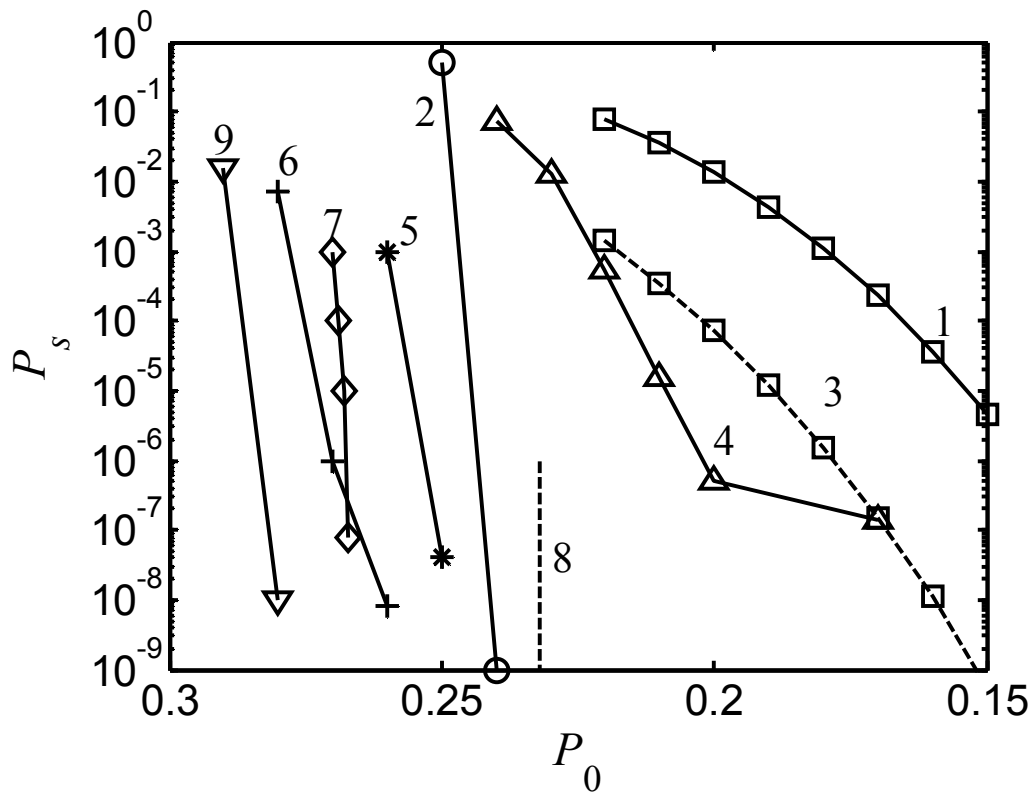


Рис. 1

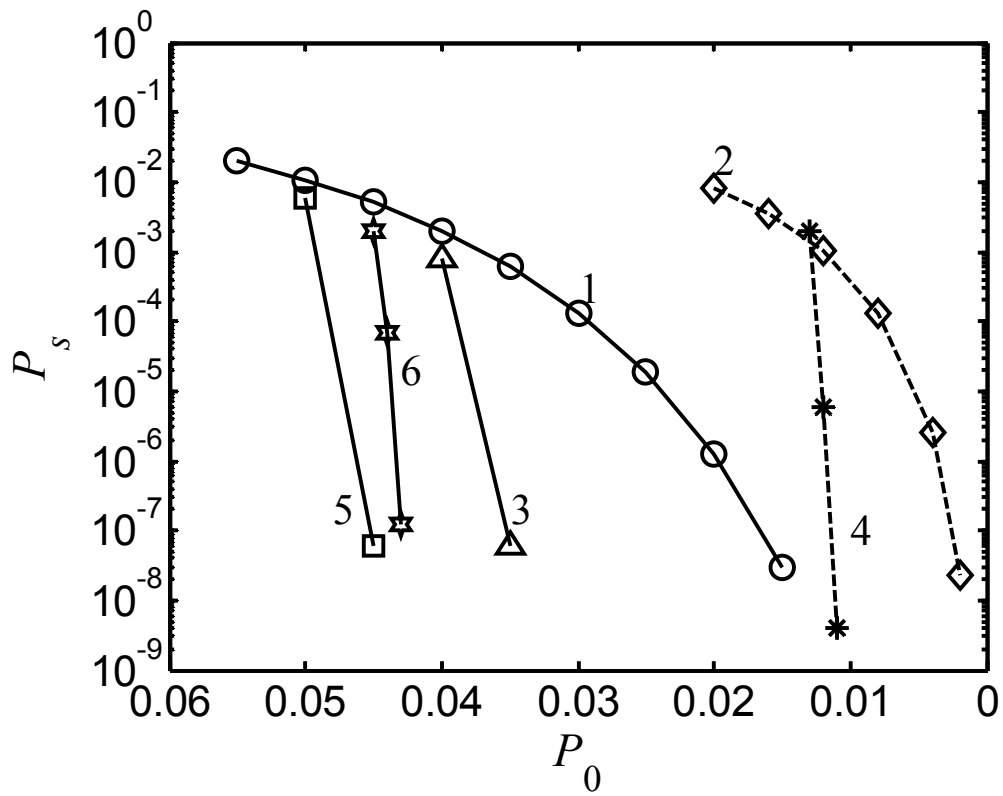


Рис. 2