УДК 004.056.3

АЛГОРИТМ ВЫБОРА КРАТНОСТИ ИСПРАВЛЯЕМЫХ ИСКАЖЕНИЙ ДЛЯ КОДИРОВАНИЯ ИНФОРМАЦИИ С ПРИМЕНЕНИЕМ КОДОВ РИДА-СОЛОМОНА

Рахман П.А.

ФГБОУ ВПО «Уфимский государственный нефтяной технический университет», Филиал в г. Стерлитамаке, e-mail: pavelar@yandex.ru

В данной статье рассматривается вероятностный анализ невосстанавливаемых искажений кадров данных при передаче по сети, закодированных с применением кодов Рида-Соломона. Также рассматривается анализ вероятностей искажения байта, кадра данных, состоящего из заданного числа байтов, и алгоритм выбора кратности исправляемых искажений при заданной вероятности искажения бита, размере кадра, длине поля полезной информации и допустимой вероятности невосстанавливаемого искажения кадра. Приводится пример оценки целесообразности применения кодов Рида-Соломона и выбора кратности исправляемых искажений.

Ключевые слова: передача данных, информационное резервирование, исправление искажений, коды Рида-Соломона

SELECTION ALGORITHM FOR THE MULTIPLICITY OF CORRECTABLE ERRORS FOR INFORMATION CODING WITH APPLICATION OF REED-SOLOMON CODES

Rahman P.A.

Ufa State Petroleum Technological University, Sterlitamak branch, e-mail: pavelar@yandex.ru

This paper deals with the probability analysis of unrecoverable errors in data frames, encoded with application of Reed-Solomon codes and transmitted over network. Error probability analysis for a single byte and data frame with a given number of bytes, and selection algorithm for multiplicity of correctable errors for given bit error probability, length of data frame, length of user data and upper limit for probability of unrecoverable error in data frame are also overviewed. Example of reasonableness assessment for application of Reed-Solomon codes and selection for the multiplicity of correctable errors are also provided.

Keywords: data transmission, information redundancy, error correction, Reed-Solomon codes

Одна из наиболее острых проблем в информационных технологиях – это защита данных от искажений. Как каналы передачи данных, так и носители информации на сегодняшний день остаются далекими от совершенства, несмотря на все усилия производителей современных аппаратных средств. Кабельные и беспроводные линии передачи информации подвержены воздействию внешних помех, искажающих форму передаваемых сигналов и тем самым делающих невозможным однозначное распознавание информации на стороне приемника, магнитные и оптические носители информации чувствительны к физическим повреждениям, делающим невозможным чтение информации из отдельных участков на поверхности носителя.

В настоящее время в системах хранения и передачи данных применяют различные технологии информационного резервирования с применением специальных алгоритмов кодирования на базе корректирующих кодов, в частности, кодов Рида-Соломона [1], которые за счет использования избыточной информации делают возможным исправление искажений. Однако, введение избыточной информации снижает долю полезной информации в передаваемых сетевых кадрах, соответственно, возникает задача анализа вероятностей [2] искажения информации, и выбора между избыточностью и устойчивостью к искажениям.

В рамках научных исследований автора в области надежности систем хранения, передачи и обработки данных [3-10] возникла научная задача вероятностного анализа невосстанавливаемых искажений кадров при передаче по каналам связи, закодированных с применением кодов Рида-Соломона, и разработки алгоритма выбора оптимальной кратности исправляемых искажений. Соответственно, автором было проведен анализ вероятностей искажений, и имитационное моделирование кодирования, искажения и декодирования сетевых кадров. Также был предложен простой алгоритм выбора кратности исправляемых искажений при заданной вероятности искажения бита, размере кадра, длине поля полезной информации и допустимой вероятности невосстанавливаемого искажения кадра.

Кодирование информации с применением кодов Рида-Соломона. В современной практике при использовании кодов Рида-Соломона чаще всего применяется

кодирование информации, представленных в виде блоков размером k байтов. Далее путем специального алгоритма кодирования, базирующегося на алгебраическом представлении информации в виде полиномов над полем Галуа GF(2^8) и их преобразования над этим полем вычисляются r контрольных байтов. При систематическом кодировании контрольные байты добавляются к информационным байтам, и в итоге получается кадр размером n байтов (рис. 1).

Рис. 1. Структура кадра с информационными и контрольными байтами

Количество контрольных байтов чаще всего выбирается четным числом, и оно равно удвоенной кратности t исправляемых искажений – максимальному количеству искаженных байтов, которые можно гарантированно восстановить, используя алгоритм декодирования.

Таким образом, $n = k + r$ и $r = 2t$. При использовании кодов Рида-Соломона для байтовых блоков, максимальный размер кадра составляет $2^8 - 1 = 255$ байтов в силу свойств поля Галуа GF(2^8). Кроме того, при фиксированной длине кадра при увеличении числа r контрольных байтов с целью увеличения кратности исправляемых искажений t и снижения вероятности невосстанавливаемого искажения кадра (когда при передаче кадра искажается более t байтов), на долю информационных байтов остается меньшее количество $n - 2t$.

Анализ вероятностей искажения информации. Информация может искажаться либо при передаче по каналам передачи данных, либо при хранении на каком-либо носителе информации (жесткий диск, оптический диск и т.п.).

Для каналов передачи данных, как правило, известна вероятность искажения бита. Данные передаются в виде последовательности битов, и каждый из них может подвергнуться искажению. Здесь мы сделаем несколько важных допущений (ради упрощения анализа):

● Вероятность искажения того или иного бита в том или ином байте в канале передачи данных одна и та же, и не зависит от позиции байта в кадре и бита внутри байта.

● Канал передачи данных не обладает «памятью», и вероятность искажения очередного передаваемого бита не зависит от того, были ли искажены предыдущие биты.

● Вероятность искажения бита не меняется со временем или меняется достаточно медленно, и в пределах отрезка времени, требуемого для передачи кадра, вероятность искажения бита можно считать постоянной.

При соблюдении вышеперечисленных условий, передачу кадра длины n можно считать последовательностью из n независимых испытаний по передаче отдельных байтов, в свою очередь, в рамках передачи отдельного байта мы имеем дело с последовательностью из 8 независимых элементарных испытаний по передаче отдельных битов. В итоге мы имеем дело с последовательностью из $8n$ элементарных независимых испытаний, которую мы можем также интерпретировать, как n независимых серий по 8 независимых элементарных испытаний в каждой серии. Особо отметим также, что в случае искажения нескольких битов, они могут различными способами располагаться внутри кадра, состоящего из n байтов, например, 8 искаженных битов могут «уместиться» внутри одного байта, а могут «распылиться» по 8 различным байтам – и это будут принципиально различные ситуации с точки зрения корректирующей способности кодов Рида-Соломона.

Чтобы оценивать вероятности искажения одного, нескольких или всех байтов в кадре на основе информации о базовой вероятности искажения одного бита мы должны обратиться к математическому аппарату теории вероятностей.

Пусть, p – заданная вероятность искажения бита. Тогда, вероятность того, что исказится байт, равна вероятности того, что хотя бы один бит в байте исказится:

$$P_{\text{byte}} = 1 - (1 - p)^8. \qquad (1)$$

Тогда, согласно биномиальному закону распределения числа искаженных байтов, получаем вероятность того, что исказится ровно h байтов в кадре, состоящего из n байтов, при заданной вероятности p искажения бита:

$$P(T = h) = C_n^h (1 - (1 - p)^8)^h (1 - p)^{8(n-h)}. \qquad (2)$$

Искаженные биты могут «попадать» в какие-то байты, а в какие-то «не попадать». Формула определяет вероятность искажения ровно h байтов, при условии

целостности остальных $n - h$ байтов, во всех подходящих вариантах искажения $\lambda = h \ldots 8h$ битов, при условии целостности остальных $8n - \lambda$ битов в кадре и условии, что в каждый из h байтов «попадет» хотя бы один искаженный бит в каждом варианте, и также учитываются все сочетания искаженных h байтов по n байтам в кадре.

Теперь, имея формулу для вероятности искажения ровно h байтов в кадре длиной n при заданной вероятности искажения бита p, нетрудно вывести формулы для вероятности отсутствия искажения ($T = 0$), вероятности восстанавливаемого искажения ($1 \le T \le t$) и вероятности невосстанавливаемого искажения ($T > t$).

Тогда, формула для вероятности отсутствия искажения:

$$P(T = 0) = (1 - p)^{8n}. \quad (3)$$

Формула для вероятности восстанавливаемого искажения:

$$P(1 \le T \le t) = \sum_{h=1}^{t} C_n^h (1 - (1 - p)^8)^h (1 - p)^{8(n-h)}. \quad (4)$$

Наконец, формула для вероятности невосстанавливаемого искажения:

$$P(T > t) = \sum_{h=t+1}^{n} C_n^h (1 - (1 - p)^8)^h (1 - p)^{8(n-h)}. \quad (5)$$

Выбор кратности исправляемых искажений. При использовании кодов Рида-Соломона за возможность исправления t искаженных байтов приходится «платить» $r = 2t$ контрольными байтами, и в условиях, когда у нас задана фиксированная длина кадра n, при увеличении кратности исправляемых искажений t на долю полезной информации остается все меньшее $k = n - r$ число байтов. В пределе, коды Рида-Соломона могут обеспечить исправление вплоть до $t = \lfloor (n-1)/2 \rfloor$ байтов, при этом, если n нечетно, $n - 1$ байтов будут контрольными, так как $r = 2t = 2 \cdot \lfloor (n-1)/2 \rfloor = n - 1$, и всего один байт останется на долю полезной информации, так как $k = n - (n - 1) = 1$.

Очевидно, что для выбора «разумной» кратности исправляемых искажений необходимо отталкиваться от каких-то критериев и ограничений.

Пусть заданы следующие исходные параметры для задачи выбора кратности:
n – фиксированная длина кадра в байтах.
p – базовая вероятность искажения бита.
$k_{\min}$ – требуемое минимальное число байтов полезной информации в кадре.
$P_{\max}$ – допустимая вероятность невосстанавливаемого искажения кадра.

Далее, исходя из условия что, должно соблюдаться требование по допустимой вероятности невосстанавливаемого искажения кадра и требование на минимальное число байтов полезной информации, можно предложить следующую модель задачи выбора кратности исправляемых искажений:

$$\begin{cases} P(T > t) < P_{\max} \\ n - 2t \ge k_{\min} \end{cases} \Rightarrow t^* \quad (6)$$

Если, невозможно найти параметр t, удовлетворяющий обоим условиям, то, очевидно, применение кодов Рида-Соломона нецелесообразно. Кроме того, очевидно, если вероятность искажения вообще хотя бы одного байта меньше заданной допустимой границы, то есть $P(T > 0) < P_{\max}$, то применение кодов Рида-Соломона также нецелесообразно. Ниже приведена схема алгоритма для оценки целесообразности применения кодов Рида-Соломона и выбора кратности исправляемых искажений (рис. 2).

Пример выбора кратности исправляемых искажений. Задана длина кадра $n = 36$ байтов, минимальное число байтов полезной информации $k_{\min} = 32$, вероятность искажения бита $p = 5 \cdot 10^{-4}$ и допустимая вероятность невосстанавливаемого искажения кадра $P_{\max} = 0{,}001$.

Рассчитаем сначала вероятность возникновения искажения хотя бы в одном байте кадра, и тем самым, оценим оправданность использования кодов Рида-Соломона:

$$P(T > 0) = 0{,}13414$$

Очевидно, что эта вероятность значительно выше допустимой границы $P_{\max} = 0{,}001$, так что применение кодов Рида-Соломона вполне оправдано. Теперь попробуем выбрать кратность t исправляемых искажений, исходя из двух условий:

$$\begin{cases} P(T > t) < 0{,}001 \\ 36 - 2t \ge 32 \end{cases}$$

Очевидно, для заданного требуемого числа байтов полезной информации, можно рассматривать только два варианта $t = 1$

и $t = 2$. В этих вариантах вероятности невосстанавливаемого искажения кадра по формуле 5 равны: $P(T > 1) = 0,00918$ и $P(T > 2) = 0,000412$. Очевидно, что вариант $t = 2$, удовлетворяет обоим условием.

Таким образом, применение кодов Рида-Соломона целесообразно, и выбранная кратность исправляемых искажений $t = 2$.

Заключение

Таким образом, в рамках данной статьи рассмотрен проведенный автором вероятностный анализ искажений при передаче кадров данных, закодированных с применением кодов Рида-Соломона.

Также рассмотрен предложенный автором алгоритм выбора кратности исправляемых искажений при заданной вероятности искажения бита, размере кадра, длине поля полезной информации и допустимой вероятности невосстанавливаемого искажения кадра.

Наконец, приведен пример выбора кратности исправляемых искажений.

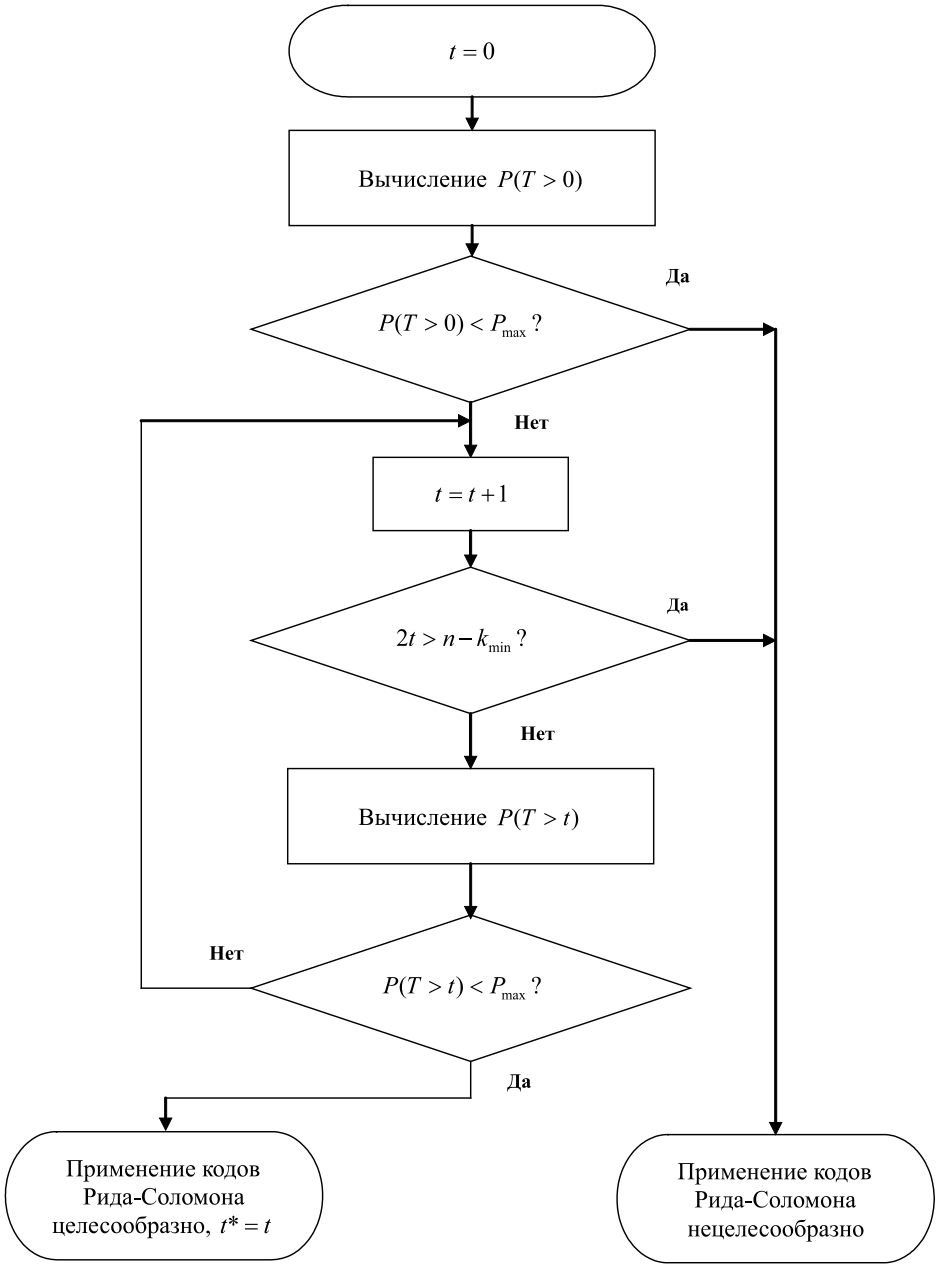

Рис. 2. Схема алгоритма для оценки целесообразности применения кодов Рида-Соломона и выбора кратности исправляемых искажений

Полученные теоретические результаты использовались в многолетней практике эксплуатации, развития и проектирования систем хранения и передачи данных НИУ МЭИ (ТУ), Балаковской АЭС, ОАО «Красный Пролетарий» и ряда других предприятий.

Список литературы

1. Todd K. Moon. Error correcting coding: mathematical methods and algorithms. – Hoboken, New Jersey: John Wiley & Sons Inc., 2005.

2. Гнеденко Б.В. Курс теории вероятностей. – М.: Едиториал УРСС, 2005.

3. Каяшев А.И., Рахман П.А., Шарипов М.И. Анализ показателей надежности избыточных дисковых массивов // Вестник УГАТУ: научный журнал УГАТУ, 2013. – Т. 17. № 2 (55). – С. 163–170.

4. Каяшев А.И., Рахман П.А., Шарипов М.И. Анализ показателей надежности локальных компьютерных сетей // Вестник УГАТУ: научный журнал УГАТУ, 2013. – Т. 17. № 5 (58). – С. 140–149.

5. Каяшев А.И., Рахман П.А., Шарипов М.И. Анализ показателей надежности двухуровневых магистральных сетей // Вестник УГАТУ: научный журнал УГАТУ, 2014. – Т. 18. № 2 (63). – С. 197–207.

6. Рахман П.А., Каяшев А.И., Шарипов М.И. Модель надежности отказоустойчивой пограничной маршрутизации с двумя Интернет-провайдерами // Вестник УГАТУ: научный журнал УГАТУ, 2015. – Т. 19. № 1 (67). – С. 131–139.

7. Рахман П.А., Каяшев А.И., Шарипов М.И. Марковская цепь гибели и размножения в моделях надежности технических систем // Вестник УГАТУ: научный журнал УГАТУ, 2015. – Т. 19. № 1 (67). – С. 140–154.

8. Рахман П.А., Каяшев А.И., Шарипов М.И. Модель надежности отказоустойчивых систем хранения данных // Вестник УГАТУ: научный журнал УГАТУ, 2015. – Т. 19. № 1 (67). – С. 155–166.

9. Рахман П.А., Шарипов М.И. Модель надежности двухузлового кластера приложений высокой готовности в системах управления предприятием // Экономика и менеджмент систем управления, 2015. – Т. 17. № 3. – С. 85–102.

10. Рахман П.А., Шарипов М.И. Модели надежности каскадных дисковых массивов в системах управления предприятием // Экономика и менеджмент систем управления, 2015. – Т. 17. № 3.1. – С. 155–168.