

УДК 681.325.3

Ю. О. Булова, ассистент (БГТУ)

КОДОВЫЕ МЕТОДЫ НЕЙТРАЛИЗАЦИИ ОШИБОК ТИПА «СТИРАНИЕ» В КАНАЛАХ ПЕРЕДАЧИ ДВОИЧНОЙ ИНФОРМАЦИИ

Статья посвящена проблеме повышения достоверности передаваемой информации в каналах с ошибками типа «стирание». Рассмотрены классические помехоустойчивые коды с фиксированной скоростью (циклические коды CRC- r и код Рида – Соломона), а также новый класс фонтанных кодов (rateless-коды): LT-код, Raptor-код, Online-код и неслучайные стирающие коды (NSSK). Проанализирована возможность применения фонтанных кодов в современных системах связи. Выявлены их преимущества и недостатки.

The article is considered the problem of reliability of transmitted information in channels with errors such as "erase". Classical redundant codes with a fixed rate (cyclic CRC- r codes and Reed-Solomon code) are reviewed. A new class of fountain codes (rateless-code): LT-code, Raptor-code, Online-code and erase codes (NSSK) are reviewed too. The using possibility of fountain codes in modern communication systems is analyzed. Their advantages and disadvantages are identified.

Введение. Все большее количество людей пользуются мобильными системами, компьютерными сетями и сетевыми приложениями, такими как цифровое телевидение через Интернет (IPTV), Интернет-телефония (VoIP), распределенные сетевые игры и др. Основными требованиями, предъявляемыми к системам связи, являются надежность и высокая скорость передачи и преобразования информации. Реальные каналы передачи данных подвергаются воздействию различных помех и шумов, многие из которых физически невозможно устранить. Это приводит к ошибкам, в результате чего возникают нежелательные свойства: часть информации может быть не доставлена до получателя, а часть – доставлена с ошибочными битами. Поэтому основной задачей теории связи является восстановление сообщения, сформированного передатчиком в одной точке пространства, в другой точке пространства спустя некоторый промежуток времени. Методы решения этой задачи зависят от свойств среды передачи информации, которые могут рассматриваться как некоторые математические модели.

Целью данной статьи является анализ существующих методов повышения достоверности передаваемой информации в каналах с ошибками типа «стирание» (erasure channel), когда решение о значении принятого символа не может считаться надежным [1].

Основная часть. В двоичном канале со стираниями передача ведется с помощью двух сигналов (0 и 1). Приемник способен каким-либо образом достоверно распознать часть принятых символов, а оставшаяся часть принятых символов может быть объявлена недостоверной или стертой, т. е. множество сигналов на выходе канала равно $\{0, 1, S\}$.

Схема переходных вероятностей канала со стираниями представлена на рис. 1.

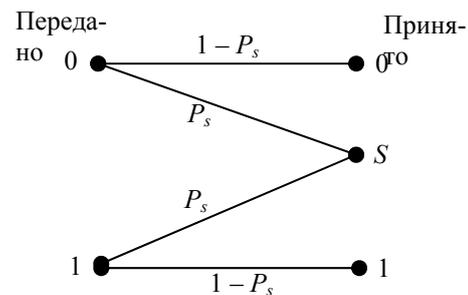


Рис. 1. Граф переходных состояний в канале с ошибками типа «стирание»

Каждому переходу двоичных символов на диаграмме приписана переходная вероятность. P_s – вероятность того, что символ будет объявлен стертым; $1 - P_s$ – вероятность достоверного приема символа. Предполагается, что вероятность ошибочного приема символа в данном канале низкая, и поэтому не учитывается.

Традиционно для повышения достоверности доставки сообщений в каналах со стиранием применяются различные методы:

- основанные на использовании обратного канала от получателя к отправителю;
- основанные на помехоустойчивом кодировании информации;
- основанные на совместном использовании корректирующего кода и перемежителя [2].

Классическими кодами, применяемыми для обнаружения и исправления одиночных и групповых ошибок в каналах передачи двоичной информации, являются циклические коды CRC- r (где $r = 4, 8, 16, 32$ – длина группы ошибок, исправляемых кодом) и код Рида – Соломона.

Однако у циклического кода CRC- r имеется ограничение: код обнаруживает все группы ошибок длиной r , но количество независимых ошибок должно быть нечетным.

Код Рида – Соломона считается оптимальным кодом для борьбы с групповыми ошибками, но имеет жесткую алгебраическую структуру, которая приводит к значительным вычислительным затратам как при кодировании, так и при декодировании. В каждый проверочный символ кода входят все исходные символы сообщения, причем число этих проверочных символов ограничено – размер закодированного сообщения не может превышать двойного размера исходного слова.

Анализ данной проблемы привел к созданию нового класса помехоустойчивых кодов для каналов со стиранием. Новый класс кодов называют классом фонтанных кодов (Digital Fountain Codes), или rateless-кодами [3]. Кодами из этого класса можно закодировать сообщение конечного размера потенциально-неограниченным потоком (фонтаном) независимых пакетов. Первый случайный фонтанный код изобрел Michael Luby в 1998 г. [3].

На рис. 2 приведен пример кодирования, используя LT-код. Пусть размер передаваемых пакетов $L = 3$. Тогда для кодирования исходного сообщения размером $K = 9$ на стороне отправителя происходит накопление трех исходных пакетов.

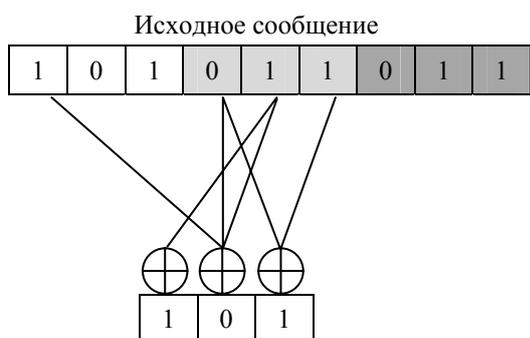


Рис. 2. Пример кодирования

Алгоритм кодирования случайного фонтанного кода происходит в три этапа:

1. Используя распределение вероятностей степеней $p(d)$, выбирается количество исходных символов d_i (степень кодового символа), вовлеченных в операцию кодирования для генерации кодового символа x_i .

2. Из сообщения, состоявшего из K исходных символов, случайным образом выбираются d_i символов.

3. Сложением бит этих исходных символов по модулю 2 (операция XOR) формируется кодовый символ x_i .

Процесс формирования кодовых символов продолжается до тех пор, пока не будет получено сообщение от приемника о достоверном приеме сообщения, либо отправитель сам принимает решение о прекращении процесса кодирования, руководствуясь принципом достаточности. Поскольку декодеру не важен порядок передачи кодовых символов и время формирования каждого пакета постоянно, кодер такого класса по запросу всегда может добавить «на лету» небольшое число кодовых пакетов [3].

Алгоритм декодирования предполагает наличие информации от кодера о степени каждого кодового символа d_i и списке номеров исходных символов. Эта информация может быть передана декодеру в заголовке сообщения либо вычислена на стороне приемника при синхронизации его с отправителем и использовании идентичных генераторов случайных чисел.

На приемнике процесс декодирования начинается с приема первого пакета и проходит в несколько этапов:

1. Находится кодовый символ x_i со степенью $d_i = 1$ (рис. 3). Если такого символа нет, декодирование прерывается на этом шаге (декодер заявляет об отказе от декодирования).

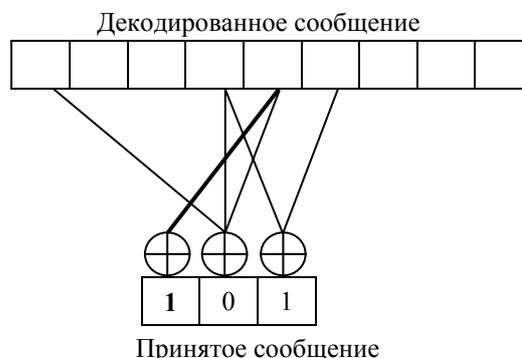


Рис. 3. Первый этап декодирования

2. Устанавливается $s_j = x_i$, тем самым находится j -й символ (рис. 4).

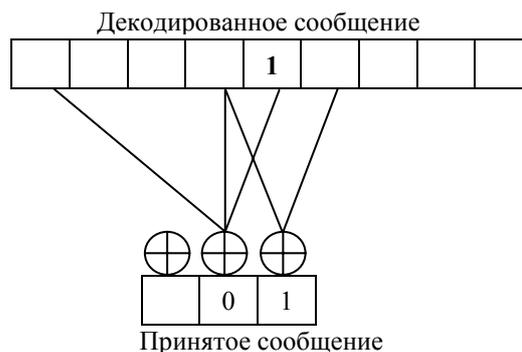


Рис. 4. Второй этап декодирования

3. Удаляются копии j -го символа из всех кодовых символов: $x_{i'} = x_i - s_j$ (рис. 5).

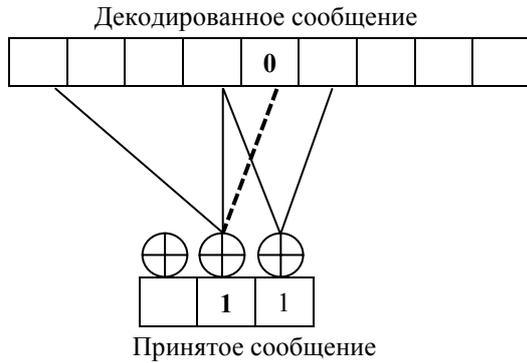


Рис. 5. Третий этап декодирования

4. Удаляются все ребра графа, идущие от найденного символа. В результате номер j удаляется из всех списков номеров кодовых символов, понижая степень соответствующих кодовых символов: $d_{i'} = d_i - 1$ (рис. 6).

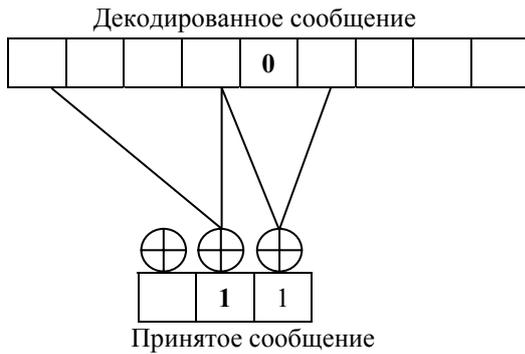


Рис. 6. Четвертый этап декодирования

5. Осуществляется переход к шагу 1, если не все K символов декодированы.

Таким образом, условие $p(1) > 0$ является обязательным для старта алгоритма декодирования. Идеальное вероятностное распределение, при котором после каждой итерации декодирования в графе оставался бы один кодовый символ степенью 1, имеет вид

$$\mu(d) = \begin{cases} \frac{1}{K} & \text{при } d = 1, \\ \frac{1}{d(d-1)} & \text{при } d = 2, 3, \dots, K. \end{cases} \quad (1)$$

Однако чтобы восстановить при стирании исходный символ, он должен входить в большинство кодированных символов, иными словами, степень каждого кодового символа должна стремиться к K . С другой стороны, большая часть символов должна иметь низкую степень для снижения вероятности отказа от декодирования.

Этим требованиям отвечает распределение:

$$\tau(d) = \begin{cases} \frac{S}{Kd} & \text{при } d = 1, 2, \dots, \frac{K}{S}, \\ \frac{S}{K} \cdot \ln\left(\frac{S}{\delta}\right) & \text{при } d = \frac{K}{S}, \\ 0 & \text{при } d > \frac{K}{S}, \end{cases} \quad (2)$$

где $S = c \cdot \ln(K/\delta) \cdot \sqrt{K}$ – число исходных символов, хоть один раз входящих в кодовые символы степени $d = K/S$; c – параметр, величина которого $0 \leq c \leq 1$; δ – вероятность неудачного декодирования.

Компромисс между использованием двух распределений разрешается путем использования параметрического распределения LT-кода (3), имеющего название робастное распределение (Robust Soliton Distribution):

$$p(d) = \frac{\mu(d) + \tau(d)}{Z}, \quad (3)$$

где Z – нормирующий множитель вероятностного распределения.

Стоимость декодирования оказывается порядка $\ln(K/\delta)$ операций XOR [3].

Произведены исследования исправляющей способности LT-кода относительно длины исходной последовательности бит (при 5% потерь данных и скорости кода $Rc = 0,7-0,95$). Результаты приведены на рис. 7. Из графически представленной зависимости видно, что наибольшая исправляющая способность кода наблюдается при длине исходной последовательности 1000 бит.

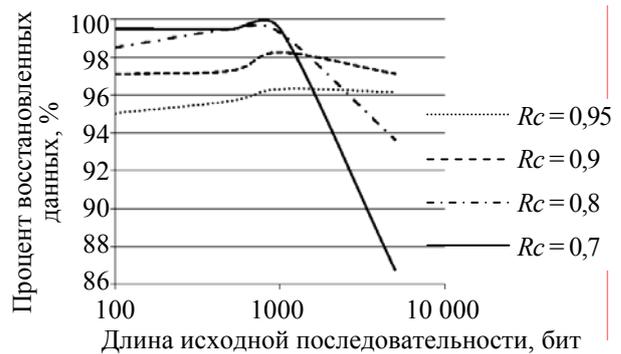


Рис. 7. Зависимость исправляющей способности LT-кода от длины исходной последовательности

Также были выполнены исследования восстанавливающей способности LT-кода относительно количества потерянных данных при длине исходной последовательности бит от 100 до 5000 бит и скорости кода $Rc = 0,9$ (рис. 8).

Из приведенной ниже зависимости видно, что при количестве стираний, превышающем 8% от исходного сообщения, при скорости кода $R_c = 0,9$ значительно уменьшается восстанавливающая способность LT-кода, поэтому целесообразно снижать скорость кода (рис. 9).

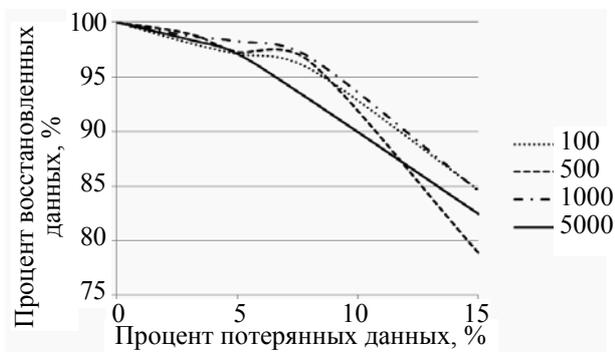


Рис. 8. Зависимость исправляющей способности LT-кода от количества потерянных данных

На рис. 9 показана зависимость способности восстановления потерянных пакетов от скорости кода при 15% потерянных данных. Таким образом, при таком количестве ошибок типа «стирание» целесообразно использовать код с избыточностью не менее 20%.

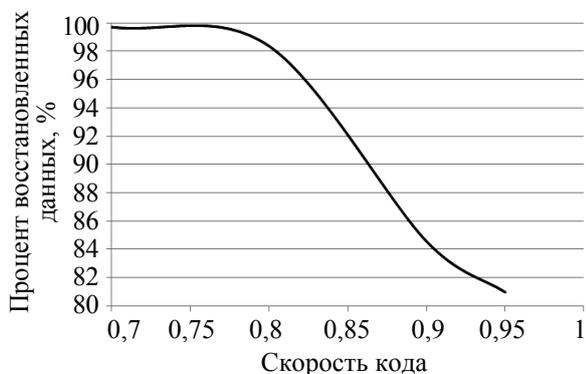


Рис. 9. Зависимость исправляющей способности LT-кода от скорости кода

В 2003 г. А. Shokrollahi разработал Raptor-код, который представляет собой каскадную конструкцию с внутренним LT-кодом и внешним блоковым кодом с фиксированной скоростью. В 2002 г. М. Маумоунков независимо от создателей LT- и Raptor-кодов разработал Online-код, который также представляет каскадную конструкцию, однако в качестве внутреннего кода используется код с отличным от LT-кода вероятностным распределением. С 2008 г. К. В. Шинкаренко и А. М. Кориков ведут разработку неслучайных стирающих кодов, которые призваны уменьшить объем потерь при IP-телевещании [4].

Алгоритмы, связанные с рассмотренными кодами, защищены большим количеством патентов и используются в таких программно-аппаратных изделиях для компьютерных сетей, как Digital Fountain Multicast Client, DF Raptorfor Streaming, DF Broadcast. Код Raptor применяется в стандарте DVB-H.

Заключение. Рассмотрены кодовые методы с фиксированной скоростью, а также gateless-коды, направленные на исправление ошибок типа «стирание» в каналах передачи двоичной информации. Классические коды показывают хорошие результаты в исправлении пакетных ошибок некоторой кратности, однако имеют сложные и ресурсоемкие алгоритмы кодирования и декодирования.

Коды, относящиеся к классу фонтанных кодов, обладают как преимуществами, так и недостатками. Они имеют простые и быстрые алгоритмы кодирования и декодирования, принципиально не зависящие от размера пакета; могут быть использованы в любом канале со стираниями, независимо от статистики стираний; позволяют на практике получать результаты, близкие к предельным возможностям помехоустойчивого кодирования, однако эффективны лишь при больших объемах данных. К недостаткам применения фонтанных кодов следует отнести неизбежность буферизации относительно большого объема данных как на передатчике, так и на приемнике; необходимость синхронизации отправителя и получателя для генерации случайных чисел либо передача дополнительной информации по каналу. Следует также учитывать, что данные коды не эффективны в каналах, где помимо ошибок типа «стирание» встречаются ошибочные приемы символов.

Литература

1. Гладких, А. А. Основы теории мягкого декодирования избыточных кодов в стирающем канале связи / А. А. Гладких. – Ульяновск: УлГТУ, 2010. – 380 с.
2. Gorbunova, Yu. W-cyclic method interleaving of data for communication system / Yu. Gorbunova, P. Urbanovich // News electrical and electronic technologies and their industrial implementation: proc. VII International Conference, Zakopane, Poland, June 28 – July 1, 2011. – Zakopane, 2011. – P. 149.
3. Варгаузин, В. Помехоустойчивое кодирование в пакетных сетях / В. Варгаузин // ТелеМультиМедиа. – 2005. – № 3. – С. 10–16.
4. Шинкаренко, К. В. Применение неслучайных стирающих кодов в IP-телевещании / К. В. Шинкаренко, А. М. Кориков // Известия Томского политехнического университета. – 2008. – Т. 313, № 5. – С. 42–46.

Поступила 28.02.2012