Материалы Международной научно-технической конференции,
21 – 25 ноября 2016 г.

МОСКВА **INTERMATIC – 2 0 1 6, часть 5** **МИРЭА**

ЭФФЕКТИВНОЕ ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ МЕТРИЧЕСКОГО ИСТОЧНИКА ДЛЯ ОБОБЩЕННОГО РАСПРЕДЕЛЕНИЯ ПОЯВЛЕНИЯ ОШИБОК КАК ФУНКЦИИ СТРУКТУРЫ КОДА

© 2016 г. А.Э. БАТАЛОВ, И.С. СИНЕВА

Московский технический университет связи и информатики
e-mail: i.alexey.batalov@gmail.com

Для эффективного помехоустойчивого кодирования ранее был предложен алгоритм генетического типа, опирающийся на структуру источника сообщений, который погружен в некоторое метрическое пространство. В то же время происхождение и структура кодовой последовательности таковы, что разные символы могут искажаться с различной вероятностью. В работе описывается обобщение генетического алгоритма-прототипа на случай не равновероятного появления искажений в разных символах кодовой последовательности.

Описание алгоритма генетического кодирования источника

В основу классического помехоустойчивого кодирования положен принцип дополнительной избыточности [1]. В каждое информационное сообщение добавляются дополнительные, проверочные символы. Но внесение дополнительной избыточности в передаваемое сообщение может оказаться неприемлемой ценой. Поэтому возникает задача уменьшения последствий возможных искажений символов при передаче без внесения дополнительной избыточности в сообщение. Для такого класса задач были разработаны генетические алгоритмы кодирования источника. Этот класс алгоритмов позволяет минимизировать вред от ошибок за счет того, что близкие сообщения в пространстве источника сообщений, также являются близкими в пространстве кодовых сообщений [2,3].

Принцип работы алгоритма-прототипа

Кодирование происходит итеративно, по шагам, где на каждом шаге выбираются точки текущей категории. Категория – количество единиц (или нулей) в кодовой комбинации [4]. Тем самым, для 2^n точек будет n категорий. Для удобства описания не будем рассматривать случаи, когда число точек не равно 2^n [5] Чуть более подробно рассмотрим первые итерации алгоритма. В самом начале выбирается точка нулевой категории, которая получит кодовою комбинацию, состоящую из всех нулей. Как правило, выбирается центр масс, но это не обязательное условие.

$$g_{0,1} = \arg\min\{\sum_{i=1}^{2^N} D(i,j)\},$$

где $D(i,j)$ - элемент матрицы расстояний D в метрике исходного пространства. Далее выбираются точки первой категории, кодовые комбинации с одной единицей. Это n наименее удаленных точек от точки нулевой категории.

$$g_{1,1} = \arg\min\{D(\mathrm{g}_{0,1}, j), 1 \le j \le 2^N, j \ne g_{0,1}\}$$

$$g_{1,2} = \arg\min\{D(\mathrm{g}_{0,1}, j), 1 \le j \le 2^N, j \ne g_{0,1}, j \ne g_{1,1}\}$$

$$g_{1,\mathrm{N}} = \arg\min\{D(\mathrm{g}_{0,1}, j), 1 \le j \le 2^N, j \ne g_{0,1}, j \ne g_{1,1}, j \ne g_{1,2}, \ldots, j \ne g_{1,\mathrm{N}-1}\}$$

При этом комбинации новым точкам присваиваются в произвольном порядке. Далее происходит поиск точек второй категории. Здесь уже берется не одна точка, а все возможные пары точек первой категории. Для каждой пары находится такая точка, которая наименее удалена от этой пары. При этом новая точка получает такую кодовую комбинацию, которая имеет единичное хэмминговое расстояние от соответствующей пары точек первой категории. Далее алгоритм повторяется, пока все точки не получат кодовые комбинации. В общем виде алгоритм представляется следующим образом:

$$g_{i+1,k} = \arg\min\{DP_{i+1}(k, j), 1 \le k \le \mathrm{C}_{2^N}^{i+1}, 1 \le j \le 2^N\},$$

где $DP_{i+1}(k, j)$ - преобразованная матрица расстояний, где каждая строка – это сумма всех возможных групп из $n-1$ точек предыдущей категории.

$$dp_{i+1,j} = \sum_{j=1}^{n-1} d_{g_{i,j}}$$

Наличие ошибки

Представленный выше алгоритм никак не учитывает распределение ошибок в различных символах двоичного кода. В то же время кодирующие устройства и канал могут приводить к тому, что вероятности появления искажений в различных битах отличаются. Учет этого обстоятельства на этапе кодирования позволит сократить последствия появления побитовых искажений с точки зрения минимизации расстояний между отправленным символом и декодированным в метрике пространства источника. Ранее было рассмотрено поведение различных алгоритмов кодирования при наличии ошибки [6 – 8] Опишем модификацию алгоритма-прототипа, учитывающую данное обстоятельство.

Выбор точки нулевой категории никак не меняется: это может быть либо центр масс, либо точка сгущения. Для поиска точек первой категории необходимо учитывать, что наиболее близкая точка к точке нулевой категории может оказаться не самой близкой, при возникновении ошибки в каком-то конкретном символе. Поэтому для поиска наименее удаленной точки необходимо учитывать возможное искажения в каждом символе, вероятность искажения символа i обозначим e_i. Если в кодовом слове длины N искажены символы $i_1, i_2, ..., i_k$, то вероятность этого события

$$\pi(i_1, i_2, ... i_k) = (1-e_1)\cdot\ldots\cdot(1-e_{i_1-1})\cdot e_{i_1}\cdot(1-e_{i_1+1})\cdot\ldots\cdot(1-e_{i_2-1})\cdot e_{i_2}\cdot(1-e_{i_2+1})\cdot\ldots\cdot$$
$$\cdot(1-e_{i_k-1})\cdot e_{i_k}\cdot(1-e_{i_k+1})\cdot\ldots\cdot(1-e_N)$$

Поскольку все вероятности e_i малы, разложение $\pi(i_1, i_2, ..., i_k)$ по малым параметрам дает главный член разложения

$$\pi(i_1, i_2, ..., i_k) = e_{i_1}\cdot e_{i_2}\cdot\ldots\cdot e_{i_k} + o\left(e_{i_1}\cdot e_{i_2}\cdot\ldots\cdot e_{i_k}\right)$$

Наиболее вероятно появление одного искажения, поэтому за основу модификации алгоритма-прототипа будет взята линеаризация вероятностей перехода одних кодовых комбинаций в другие по параметрам e_i.

Для точек первой категории находим

$$g_{1,1} = \arg\min\{D(\mathrm{g}_{0,1}, j)e_i, 1 \le j \le 2^N, 1 \le i \le N, j \ne g_{0,1}\}$$
$$g_{1,2} = \arg\min\{D(\mathrm{g}_{0,1}, j)e_i, 1 \le j \le 2^N, 1 \le i \le N, j \ne g_{0,1}, j \ne g_{1,1}\}$$
$$\ldots$$
$$g_{1,\mathrm{N}} = \arg\min\{D(\mathrm{g}_{0,1}, j)e_i, 1 \le j \le 2^N, 1 \le i \le N, j \ne g_{0,1}, j \ne g_{1,1}, j \ne g_{1,2}, \ldots, j \ne g_{1,\mathrm{N}-1}\}$$

Кодовая комбинация, которую получит точка первой категории, состоит из n-1 нулей и одной единицы. При этом позиция единицы напрямую зависит от того, искажения в каком символе наименее отдаляет данную точку от точки нулевой категории.

Для точек второй категории нужно уже учитывать возможные искажения пары символов. При этом для каждой пары точек первой категории однозначно определена точка второй категории. Тем самым, соседняя точка второй категории будет получаться из первой с вероятностью e_j, из второй – с вероятностью e_i. В таком случае, искомая точка второй категории будет решением следующего выражения:

$$g_{1,k} = arg\,min\left\{d_{ki}e_i + d_{kj}e_j, n+1 \le k \le 2^n\right\}$$

Поиск точек последующих категорий происходит аналогично.

$$g_{i+1,k} = \arg\min\{DE_{i+1}(k,j), 1 \le k \le \mathrm{C}_{2^N}^{i+1}, 1 \le j \le 2^N\},$$

где $DE_{i+1}(k,j)$ - модифицированная промежуточная матрица суммы расстояний, где каждая строка получается следующим образом:

$$de_{i+1,j} = \sum_{j=1}^{n-1} d_{g_{i,j}} e_j$$

Тем самым, такая модификация алгоритма позволит более корректно учесть неравномерную природу возникновения ошибок при кодировании и передаче сигнала.

Сравнение алгоритмов кодирования

Проведем сравнение модифицированного генетического алгоритма кодирования и случайного кодирования. Для этого возьмем пространство источника сообщений, которое хорошо описывает нормальное распределение N(0,1) и имеющего 1024 точки. Рассмотрим для примера экспоненциально убывающие битовые ошибки вида $p_i = i10^{-2i}$. Проведем 1000 экспериментов и сравним, на какое расстояние в среднем отклоняются сообщения, закодированные генетическим алгоритмом и случайным. Получились следующие результаты: для случайного кодирования среднее расстояние между исходным и искаженным сообщением составляет 1.83, а для генетического алгоритма кодирования: 0.72. В относительных единицах средняя ошибка для генетического алгоритма отстоит от средней ошибки случайного на 1.2 сигма.

Примеры кодирования для различных типов ошибок

Далее рассмотрим, как изменяется результат работы алгоритма кодирования от различных типов ошибок.

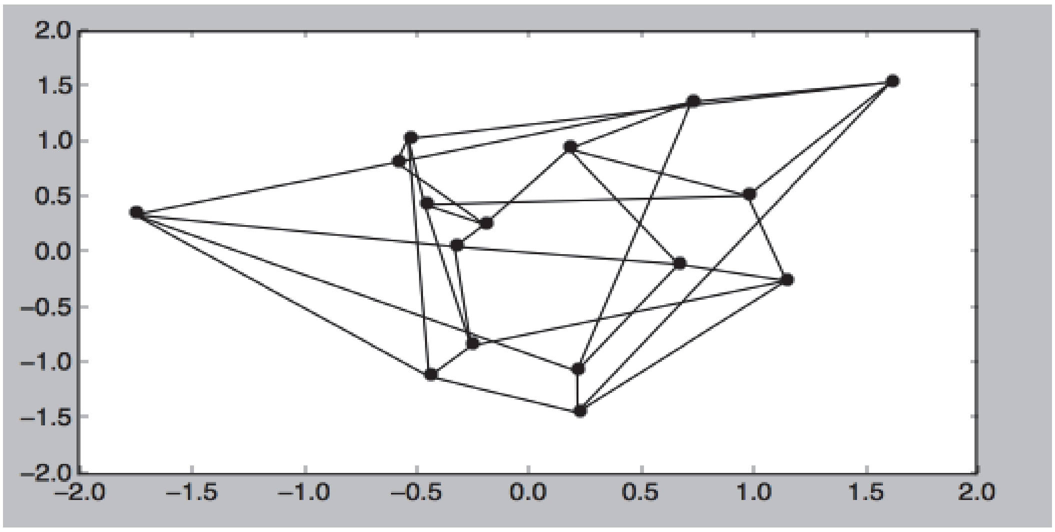

Рис. 1. Граф соседства кодовых комбинаций, полученный генетическим алгоритмом-прототипом без учета различий вероятностей ошибки.

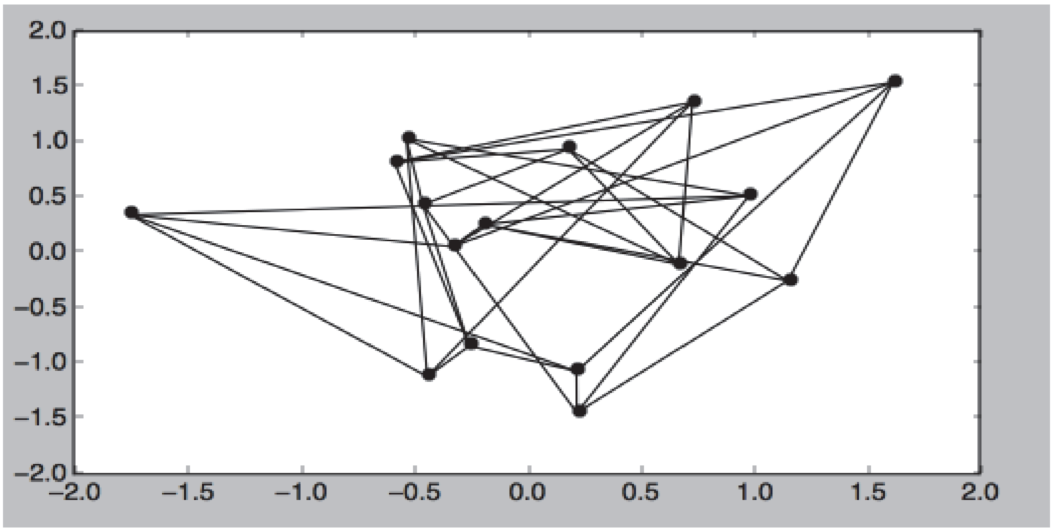

Рис. 2. Граф соседства кодовых комбинаций, полученный произвольной расстановкой без учета различий вероятностей ошибки.

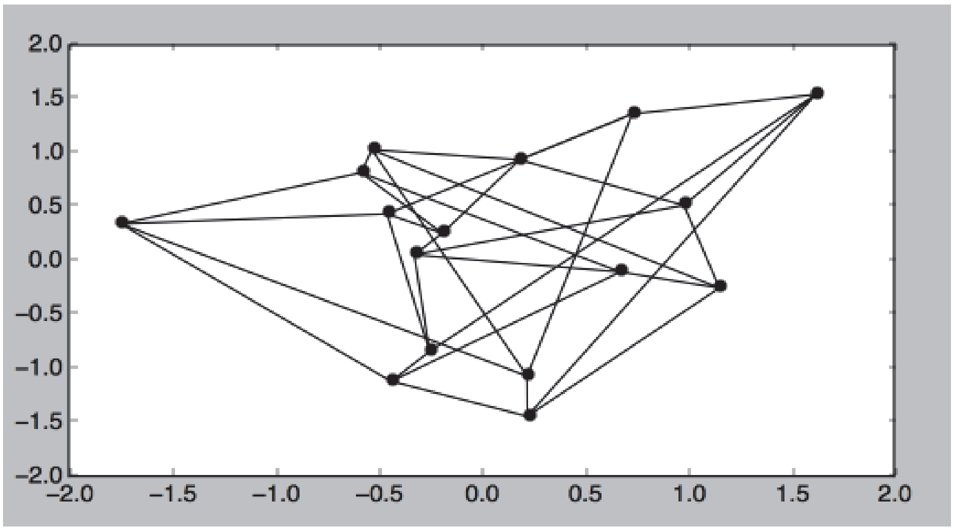

Рис. 3. Граф соседства кодовых комбинаций, полученный модифицированным генетическим алгоритмом при линейных вероятностях битовых ошибок.

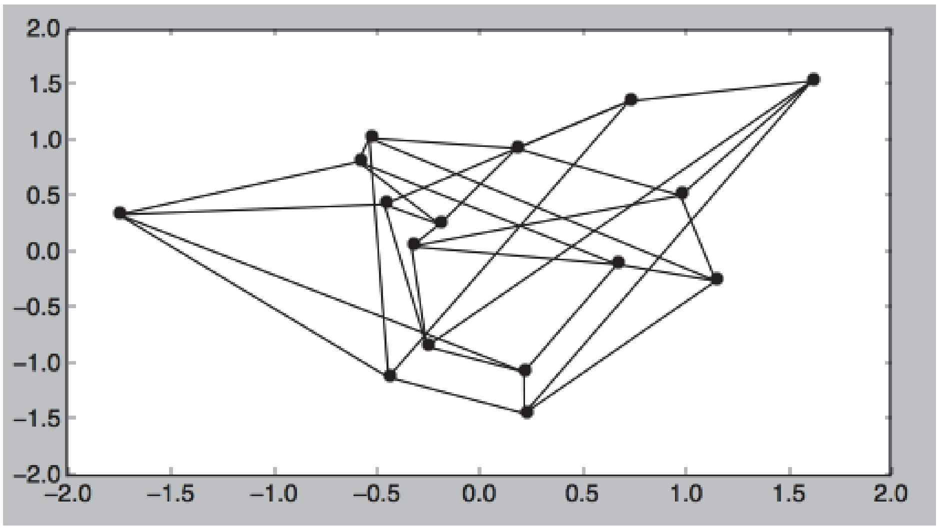

Рис. 4. Граф соседства кодовых комбинаций, полученный модифицированным генетическим алгоритмом при экспоненциальных вероятностях битовых ошибок.

На рис. 1 и 2 показаны результаты работы генетического алгоритма и случайного кодирования без учета ошибок. Далее результат случайного кодирования приводиться не будет, т.к. случайное кодирование не учитывает характер ошибок в канале.

На рис. 3 представлен результат работы для ошибок, которые выражаются таким отношением: $p_i = i10^{-2i}$. На рис. 4 ошибки экспоненциально затухают: $p_i = 0.01e^{(-0.5i)}$

Заключение

Предложенная модификация алгоритма генетического кодирования позволит учитывать неравномерную природу появления ошибок в канале. Было показано, что данный алгоритм кодирования дает существенный выигрыш по сравнению со случайными алгоритмами кодирования

СПИСОК ЛИТЕРАТУРЫ

1. *Морелос-Сарагоса Р.* Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение, Москва: Техносфера, 2006. – 320 с.
2. *Фенчук М.М., Синева И.С.* Анализ помехоустойчивости генетического кодирования с применением циклического избыточного кода // T-Comm: Телекоммуникации и транспорт. – 2014. – №11. – С. 108-112.
3. *Баталов А.Э., Синева И.С.* Сравнительный анализ помехоустойчивых свойств генетических алгоритмов безызбыточного кодирования для кластеризующихся пространств источника // T-Comm: Телекоммуникации и транспорт. – 2015. – №1. – С. 68-74.
4. *Баталов А.Э., Синева И.С.* Оптимизация алгоритма генетического кодирования источника сообщений // T-Comm: Телекоммуникации и транспорт. – 2014. – №12. – С. 6-9.
5. *Фенчук М.М., Синева И.С.* Оптимизация алгоритма генетического кодирования для пространств произвольных размерностей // T-Comm: Телекоммуникации и транспорт. – 2015. – Том 9. – №7. – С. 74-79.
6. *Фенчук М.М., Баталов А.Э., Синева И.С.* Сравнительная помехоустойчивость кодов Грея и алгоритмов генетического типа // Фундаментальные проблемы радиоэлектронного приборостроения. – М.: Энергоатомиздат, 2014, часть 5. – С. 44-47.
7. *Синева И.С., Баталов А.Э.* Повышение устойчивости совершенного кода Хэмминга к воздействию импульсных помех с использованием генетического кодирования источника // Фундаментальные проблемы радиоэлектронного приборостроения. – М.: Энергоатомиздат, 2013, часть 4. – С. 150-154.
8. *Фенчук М.М, Баталов А.Э., Синева И.С.* Повышение помехоустойчивости кодов CRC при помощи предварительного генетического кодирования метризованного источника сообщений // Фундаментальные проблемы радиоэлектронного приборостроения. – М.: Энергоатомиздат, 2013, часть 4. – С. 65-70.