РУХАНИ ЖАҢҒЫРУ

20 АСТАНА

ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ

Студенттер мен жас ғалымдардың
«ҒЫЛЫМ ЖӘНЕ БІЛІМ - 2018»
XIII Халықаралық ғылыми конференциясы

СБОРНИК МАТЕРИАЛОВ

XIII Международная научная конференция
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ - 2018»

The XIII International Scientific Conference
for Students and Young Scientists
«SCIENCE AND EDUCATION - 2018»

12th April 2018, Astana

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ

Студенттер мен жас ғалымдардың «Ғылым және білім - 2018» атты XIII Халықаралық ғылыми конференциясының БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ XIII Международной научной конференции студентов и молодых ученых «Наука и образование - 2018»

PROCEEDINGS of the XIII International Scientific Conference for students and young scholars «Science and education - 2018»

2018 жыл 12 сәуір

Астана

УДК 378

ББК 74.58
Ғ 96

Ғ 96
«Ғылым және білім – 2018» атты студенттер мен жас ғалымдардың XIII Халықаралық ғылыми конференциясы = XIII Международная научная конференция студентов и молодых ученых «Наука и образование - 2018» = The XIII International Scientific Conference for students and young scholars «Science and education - 2018». – Астана: http://www.enu.kz/ru/nauka/nauka-i-obrazovanie/, 2018. – 7513 стр. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-997-6

Жинаққа студенттердің, магистранттардың, докоранттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 378

ББК 74.58

ISBN 978-9965-31-997-6

толықтырылатын құрал ретінде шығарылады. Тіл C++ бағдарламаханалары арқылы толықтырылуы мүмкін. Техникалық нюанстарды түсінгісі келген пайдаланушылар C++ негізделген AVR-C тіліне ауыса алады. Тиісінше, AVR-C ортасынан кодты Arduino бағдарламасына қосуға болады. Arduino компаниясының негізі ATMEGA8 және ATMEGA168 микроконтроллерлерінің сұлба диаграммаларын кеңейту және ашу мүмкіндігімен жабдықталған. Модульдік сұлбалар Creative Commons лицензиясымен беріледі, яғни тәжірибелі инженерлер модульдердің өз нұсқаларын жасауға, оларды кеңейтуге және толықтыруға мүмкіндік алады. Тіпті қарапайым пайдаланушылар ақшаны үнемдеу және жұмысын түсіну үшін прототиптер әзірлеуі мүмкін. Бұл ғылыми жоба жалпы айтқанда Arduino туралы әңгімелейді, даму ортасы, даму бағдарламалары бар бірнеше бағдарлама, код жазу, Arduino Uno үшін кодты құлыптаудың графикалық көрінісі.

Қолданылған әдебиеттер тізімі:

1. Әубәкір Д.Ә. Жүйелер теориясының негіздері. Основания теории систем. Bases of Systems Theory. Оқулық/ Учебник/ Textbook. Астана: Л.Н.Гумилев ат-ғы ЕҰУ баспасы, 2011. – 500 б.

2. Канаков В.А. Новые технологии измерения в цифровых каналах передачи информации. Учебно-методический материал по программе повышения квалификации «Современные системы мобильной цифровой связи, проблемы помехозащищенности и защиты информации». – Нижний Новгород, 2006. – 91 с.

3. Brian W. Evans. Arduino блокнот программиста.

4. Абдуллаев Д.А., Арипов М.Н. Передача дискретных сообщений в задачах и упражнениях. – М.: Радио и связь, 1985. Зограф Ф.Г. Основы компьютерного проектирования и моделирования РЭС (в ПП OrCAD). Лабораторный практикум. – Красноярск: Сиб. Фед. Ун-тет, 2011. – 120 с.

УДК 004.032.97

СЖАТИЕ. КОДИРОВАНИЕ И ШИФРОВАНИЕ ИНФОРМАЦИИ

Рысмахан Жұмагүл
Студент ЕНУ им. Л.Н. Гумилева, Астана
Научный руководитель – У. Кабылбекова

Энтропия и неопределенность в криптосистеме

звестно, что имеется нижнее предельное значение E_b/N_0, при котором ни при какой скорости передачи нельзя осуществить безошибочную передачу информацию и тут требуется определить необходимый объем информации. Соответственно разрабатываются система связи с определенной способностью к обработке сообщений, которому нужна метрика измерения объема передаваемой информации. Шеннон вел метрику H, называемой энтропией источника сообщений (имеющего n возможных выходных значений). Энтропия определяется как среднее количество информации, приходящиеся на один выход источника, и выражается следующим образом:

$$H(X) = -\sum_{j=1}^{N} p_j \, log_2(p_j) \quad (1)$$

где H – среднее число бит на событие.

Таким образом, энтропия источника - это средний объем неопределенности, которая может быть разрешена с использованием алфавита, а также она представляет среднее количество информации, которое должно быть отправлено через канал связи для разрешения этой неопределенности [7, 8].

Следует отметить, что это количество информации в битах на символ (в виде энергии битов на символ), ограниченным снизу нулем (минимальная энергия), если не

существует неопределенности, и сверху $\log_2$(N), если неопределенность максимальна [5, 6]

$$0 \leq H(X) \leq log_2(\mathrm{N}), \quad (2)$$

Следует отметить, что для информационного источника с двумя равновероятным состоянием (например, выбрасывание монеты правильной формы), неопределенность исхода известен, следовательно, среднее количество информации максимальны, как только вероятности уходят от равновероятного состояния, среднее количество информации снижается. В пределе, когда одна из вероятностей обращается в нуль, H также обращается в нуль. Результат известен до того, как произойдет событие, так что исход не несет в себе дополнительной информации. Например, событие рождения ребенка (отсутствие информации), а вот рождение мальчика или девочки, это уже наличие информации в сообщении. Таким образом, объем информации в сообщении связан с вероятностью появления сообщения.

Если все сообщения равновероятны , мы неможем быть уверенным о возможности предсказания появления конкретного сообщения, и неопределенность информационного содержания сообщения является максимальной. Значит энтропия $H(K)$ определяется как средний объем информации на сообщение. Она может рассматриваться как мера того, насколько в выбор сообщения X вовлечен случай. Оно записывается как следующее суммирование по всем возможным сообщениям

$$H(X) = -\sum_X P(X)\, log_2 P(X) = \sum_X P(X) \log_2 \frac{1}{P(X)} \quad (3)$$

Неопределенность (условная энтропия) – мера взлома шифра.

Вероятность события X при условии, что событие Y произошло и если как выше, логарифм берется по основанию 2, то $H(X)$ представляет собой *математическое ожидание числа битов* в *оптимально закодированном* сообщении *X*. Это все еще не та мера, которую хотел бы иметь криптоаналитик. Им будут перехвачены некоторые шифрованные тексты, и он захочет узнать, насколько достоверно он может предсказать сообщение (или ключ) при условии, что был отправлен именно этот конкретный шифрованный текс. Неопределенность, определенная как условная энтропия *X* при данном *Y*, является для криптоаналитика более полезной мерой при попытке взлома шифра. Она задается с помощью следующей формулы:

$$H(X|Y) = -\sum_{X,Y} P(X,Y) log_2 P(X,Y) = \sum_Y P(Y) \sum_X P(X|Y) log_2 \frac{1}{P(X|Y)}. \quad (4)$$

Неопределенность может рассматриваться как неуверенность в том, что отправлено было сообщение X, при условии получения Y. Желательным для крипто аналитика является приближение H(X | Y) к нулю при увеличении объема перехваченного шифрованного текста Y.

Для информационного источника с двумя выше приведенными обстоятельствами говорит о том, что повышение производительности источника возможно не только за счет изменения энтропии, но и за **счет снижения средней длительности формирования знака (символа).**

2.Сжатие и кодирование

Сжатие

Рассмотрим выборочное множество сообщений, состоящее из восьми равновероятных сообщений $\{X\} = X_1, X_2, \dots, X_8$.

а) Найти энтропию, связанную с сообщением из множества $\{X\}$.

б) Дано другое множество равновероятных сообщений $\{Y\} = Y_1, Y_2$.

Пусть появление каждого сообщения Y сужает возможный выбор X следующим образом:

-- При наличии Y_1 возможны только X_1, X_2, X_3, X_4

– При наличии Y_2 возможны только X_5, X_6, X_7, X_8

Найти неопределенность сообщения X, обусловленную сообщением Y.

Решение

а) $P(X) = \frac{1}{8}$

$$H(X) = 8\left[\frac{1}{8}\log_2 8\right] = 3 \text{ бит/сообщение}$$

б) $P(Y) = \frac{1}{2}$. Для каждого Y, $P(X|Y) = \frac{1}{4}$ для четырех сообщений из множества $\{X\}$ и $P(X|Y) = 0$ для оставшихся четырех. Используя уравнение (14.6), получим следующее.

$$H(X|Y) = 2\left[\left(\frac{1}{2}\right)4\left(\frac{1}{4}\log_2 4\right)\right] = 2 \text{ бит/сообщение} \quad (5)$$

Видно, что значение Y сводит неопределенность X с 3 бит/сообщение до 2 бит/сообщений

Таким образом, с помощью понятия энтропии теория информации показывает, как вычислять вероятности строк символов алфавита, и предсказывает ее наилучшее сжатие, то есть, наименьшее, в среднем, число бит, необходимое для представления этой строки символов.

Кодирование и принцип шифрования

Известно, что при расшифровке многих систем шифрования может применяться статический анализ, основанной на статическом кодировании, где последовательность символов некоторого алфавита обладает статическими свойствами- вероятностью (частотой появления символов в последовательности). Примером может служить **арифметическое кодирование** использующий частоту появления отдельных символов и их комбинации. На этой основе рассмотрим две концепции Шеннона, усложняющие задачу криптоаналитика:

- Смешение сложная взаимосвязь между ключом и шифрованным текстом – это применение статического анализа, сужающего поиск практического подмножества областей ключей.

- Диффузия – сглаживает статистического различия между символами и их комбинациями.

Известно, что эффективное кодирование сообщений для передачи их по дискретному каналу без помех базируется на теореме Шеннона, что среднее число символов на знак сообщения $I_{\text{ср}}$ будет сколь угодно близко к величине

$$\frac{H(a)}{Log(m)}, \quad (6)$$

Тем самым реализуется свойство физических систем переходить в равновесное состояние, при котором энергия системы минимальна т.е. не существует неопределенности.

Следует отметить, что средний объем неопределенности (энтропия источника), может быть разрешен с использованием алфавита, тогда среднее количество информации, которое должно быть отправлено через канал связи для разрешения этой неопределенности, должен быть представлен в виде энергии битов на символ, ограниченным снизу нулем (минимальная энергия), если не существует неопределенности, и сверху log_2(N), если неопределенность максимальна [6],

$$0 \leq H(X) \leq log_2(\text{N}) \quad (7)$$

Информационное содержание N-символьного алфавита, используемое в действительных системах связи, обычно меньше верхнего предела соотношения (7). При использовании соответствующего кодирования источник может быть описан с помощью менее половины бита на символ (это может быть все точки между 0 и 1, причем в суперпозиции), а не одного бита на символ, как в текущей форме, в этом смысле это также нижняя граница, которая может быть достигнута с помощью некоторых кодов сжатия

данных, имеющих переменную длину. Действительный код может не достигать граничной энтропии входного алфавита, что объясняется множеством причин [6].

Для примера рассмотрим построение такой дроби на интервале [0,1], причем 0 – включая в интервал, а 1 – исключая. Разобьем интервал на под интервалы с длинами, соответствующими вероятностям появления символов в тексте. Рассмотрим, как сжимается текст в соответствии с процедурами данного алгоритма, выбрав случайный отрывок текста: «Мое молоко». Распишем вероятности появления каждого символа в тексте в порядке убывания и соответствующие символам поддиапазоны в виде табл.1.

Таблица 1.

Символ	Частота	Вероятность	Диапазон
О	4	0.4	(0,0; 0.4)
М	2	0.2	(0,4; 0.6)
Е	1	0.1	(0,6; 0.7)
Л	1	0 .1	(0,7; 0.8)
К	1	0,1	(0,8; 0,9
«_»	1	0.1	(0,9; 1.0)

Используя исходную таблицу диапазонов, кодируем текст. Исходный рабочий интервал [0,1]

Окончательная длина интервала равна произведению вероятностей всех встретившихся символов, а его начала зависит от порядка следования символов в потоке.

В качестве примера рассмотрим последовательности слово «МОЕ»

В этом примере четырем символам будут соответствовать подынтервалы (поддиапазоны -четвертый столбец таблицы 1.

Чтобы закодировать слово «МОЕ», следует начинать с интервала [0,1). Первый символ «М» сокращает интервал, отбросив от него 40% в начале и 10% в конце. Результатом будет интервал [0.4,0.6). Второй символ «О» сокращает интервал [0.4,0.6) до интервала[0.2800; 0.2400). Третий символ «Е» переводит его в [0.1600, 0.17О0). Наконец, символ «_» отбрасывает от него 90% в начале, а конечную точку оставляет без изменения и при этом получается интервал [0. 19ОО; 0.2000). Окончательным кодом нашего метода может служить любое число из этого промежутка, например - 0. 1950. Этого числа достаточно для восстановления исходной цепочки если известна исходная таблица диапазонов и длины цепочки.

(Заметим, что подынтервал [0.2800; 0.2400) получен из [0.4, 0.6) с помощью следующих преобразований его концов:

$$0.2 + (0.4 - 0.6) \times 0.4 = 0.2800 \text{ и } 0.2 + (0.4 - 0.6) \times 0.6 = 0.3200). \quad (8)$$

Таким же способом получаем и интервал [0.1600,0.1720)

На этом примере легко понять следующие шаги алгоритма арифметического кодирования:

1. Задать «текущий интервал» [0,1).

2. Повторить следующие действия для каждого символа s входного файла.

2.1. Разделить текущий интервал на части пропорционально вероятностям каждого символа.

2.2. Выбрать подынтервал, соответствующий символу s, и назначить его новым текущим интервалом.

3. Когда весь входной файл будет обработан, выходом алгоритма объявляется любая точка, которая однозначно определяет текущий интервал (то есть, любая точка внутри этого интервала). В данном случае - 0,1650. После каждого обработанного символа текущий

интервал становитсявсе меньше, поэтому требуется все больше бит, чтобы выразить его, однако окончательным выходом алгоритма является единственное число, которое не является объединением индивидуальных кодов последовательности входных символов. Среднюю длину кода можно найти, разделив размер выхода (в битах) на размер входа (в символах). Отметим, что вероятности, которые использовались на шаге 2.1, могут каждый раз меняться, и это можно использовать в адаптивной вероятностной модели [1, 12].

Декодирование Декодеру необходимо иметь либо таблицу кодов. Из таблицы кодирования (1) можно заметить, что каждый следующий интервал вложен в предыдущий, это говорит о том, что выбранная точка, число - 0,1650, соответствует первому символу «М» в цепочке , потому что только его диапазон включает данное число. В качестве интервала берется диапазон «М» - [0.4, 0.6) и в нем находиться диапазон включающий 0,1650. Перебирая всех возможных символов по приведенной таблице 1, находим, что только интервал [0.1600, 0.17O0). соответствующий диапазону для «О», включает число 0,1650. Этот интервал выбирается в качестве следующего рабочего и т.д.

Прицип генерирование ключей
Соответствие кодовых слов и сообщений

Вектор сообщения	Кодовое слово
000	000000
100	110100
010	011010
110	101110
001	101001
101	011101
011	110011
111	000111

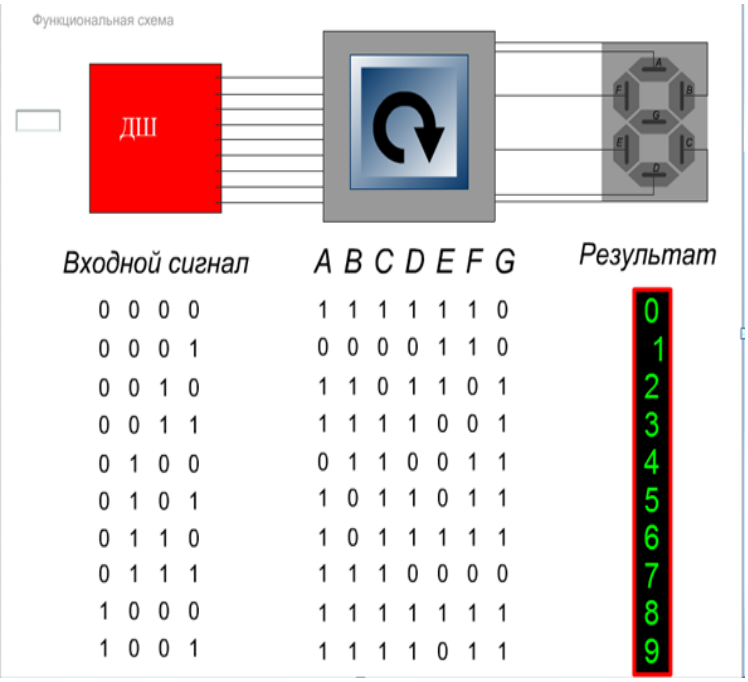
$$G = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$U_4 = [1 \quad 1 \quad 0] \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = 1 \cdot V_1 + 1 \cdot V_2 + 0 \cdot V_3 = 110100 + 011010 + 000000 =$$

101110 (кодовое слово для вектора сообщения 110)

Рисунок 5 - Сеть Bitcoina. Ключи и подписи

Криптографический алгоритм расчитывает цифровую подпись



Закрытый ключ – Пин-код, открытый ключ- номер счета

Известно, что прежде чем внести запись о транзакции, банк удостоверяет личность плательщика по паспорту , ПИН-коду или временной пароль , который высылается по SMS.

Список использованных источников

1. Сергеенко В.С, Баринов В.В. Сжатие данных, речи, звука и изображений в телекоммуникационных системах. Издательское предприятие РадиоСофт . Москва 2014. с.35 -85

2. Дмитрий Стародубцов. Технологии/ Электронные деньги. Правило Блокчейна. Ж-л. Популярная механика., М. №3. 2017 с.45-50.

3. А. Понятов. Квантовые точки прогресса. Ж-л. «Наука и жизнь №6,2016. Стр.38-45

4. Бернард Скляр. Цифровая связь. Теоретические основы и практическое применение. Москва, Санкт-Петербург, Киев:-2007с.800 - 826

5. Аифичер Э., Джервис Б.(Москва.Санкт-Петербург.Киев 2004) Цифровая обработка сигналов., Практический подход. С. 173-182

6. Дмитриев В.И. Прикладная теория информации. М.: Высшая школа, 1989. Стр.185, Игнатов В.А.Теория информации и передачи сигналов. М.: Радио и связь,1991

7. Костров Б.В. Основы цифровой передачи и кодирования информации – М.: « ТехБук», 2007. с.40 - 110

8. Лайонс Ричард. Цифровая обработка сигналов. Издательство Бином. Пер с англ. – М.: 000 Бином - Пресс», 2013. с. 156 ил

9. Конопелько В.К , Борискевич А.А, Цветков В.Ю.Многомерные технологии сжатия, защиты и коммутации изображений.Минск «Бестпринт».: -2008 с.7 – 20

10. Сэломон. Сжатие данных, изображений и звука Москва: Техносфера, 2004. - 368с. ISBN 5-94836-027-X

УДК 37.0.1