УДК 004.62

Модели векторного представления многоразрядных двоичных данных на основе псевдорегулярных чисел

Воробьев Е. Г., Хомоненко А. Д.

Постановка задачи: *получившие наибольшее распространение методы сжатия информации без потерь (кодирование длин кодов – Run Length Encoding, Хаффмана (обычный и динамический), арифметическое сжатие, словарные методы сжатия), как правило, используют статистические свойства отдельных байтов или битов текста или изображения. Наиболее распространенные алгоритмы сжатия без потерь основаны на использовании кодов переменной длины. При этом сжатие достигается путем того, что короткие коды назначаются часто встречающимся элементам данных, а длинные коды – редко встречающимся элементам. Существенным ограничением такого подхода является относительно небольшая степень сжатия данных, для которых не допускается наличие потерь. Операция сжатия является элементарной криптографической операцией.* ***Целью работы*** *является разработка математических операций для криптографических примитивов на основе модели векторного представления многоразрядных двоичных данных с использованием псевдорегулярных чисел.* ***Новизна:*** *предлагаемого подхода заключается в применении свойств двоичных чисел, связанных с их математической и структурной зависимостью от псевдорегулярных чисел для получения их короткой записи.* ***Результат****: состоит в том, что при этом увеличивается степень сжатия двоичных данных большого размера и повышается информационная безопасность их передачи и хранения. Кроме того, предложена математическая модель и описание алгоритма приведения многоразрядных двоичных чисел к псевдорегулярной структуре на основе двойного двоично-десятичного преобразования. Показано распределение чисел с псевдорегулярной структурой в упорядоченных числовых полях.* ***Практическая значимость****: представленное решение реализовано в виде демонстрационного прототипа программного приложения, реализующего сжатие, хранение и восстановление (декодирование) двоичных данных большого размера с анализом зависимости от псевдорегулярной структуры на ограниченном числовом поле и трех основных типах натуральных чисел. Предлагаемое решение может найти практическое применение с целью резервного копирования данных для повышения информационной безопасности и устойчивости перспективных информационно-вычислительных систем, функционирующих в условиях воздействий.*

Ключевые слова: *математические преобразования для криптографических примитивов, векторное представление двоичных данных, псевдорегулярные числа, хранение и восстановление данных, информационная безопасность.*

Введение

Актуальность задачи исследования обусловлена необходимостью создания современных информационных систем, реализующих единое информационное пространство. Резервирование данных в едином информационном пространстве должно опираться на современные математические методы сжатия данных с минимумом потерь, являющиеся универсальными для любого формата представления информации.

Библиографическая ссылка на статью:
Воробьев Е. Г., Хомоненко А. Д. Модели векторного представления многоразрядных двоичных данных на основе псевдорегулярных чисел // Системы управления, связи и безопасности. 2019. № 2. С. 291-303. DOI: 10.24411/2410-9916-2019-10214.

Reference for citation:
Vorobiev E. G., Khomonenko A. D. Vector Representation Models of Multi-Bit Binary Data Based on Pseudo-Regular Numbers. *Systems of Control, Communication and Security*, 2019, no. 2, pp. 291-303. DOI: 10.24411/2410-9916-2019-10214 (in Russian).

Современные методы повышения устойчивости функционирования перспективных информационно-вычислительных систем основаны на использовании архивов с высокой степенью сжатия. Используемые в настоящее подходы к сжатию текстовой информации без потерь не позволяют создавать архивы требуемого объема (с высокой степенью сжатия). Это ограничение определяется коэффициентом сжатия широко используемых методов сжатия без потерь и архиваторов, не превышающим 30, зависимостью от форматов представления и опорой на предложенную К. Шенноном в 1947 г. модель избыточности. Эта модель не обеспечивает нужную эффективность сжатия – в связи с резким ростом объемов накапливаемой человечеством информации в настоящее время требуется изменение математических принципов сжатия информации [1].

Существующие методы сжатия, лежащие в основе работы средств архивации, связи и криптографии не позволяют получить на выходе короткий двоичный код, пригодный для хранения больших объемов информации на носителях небольшой емкости. Допускается сжатие информации с потерями и зависимость от форматов представления информации.

В исследовании методов сжатия информации можно отметить следующие направления: использование методов сжатия данных без потерь при жестких ограничениях на ресурсы устройства-декодера [2]; оценка качества обучающих множеств искусственных нейронных сетей в задачах сжатия данных без потерь [3]; сжатие многоразрядных цифровых изображений без потерь на основе представления их марковскими случайными полями [4]; сжатие без потерь разностных целочисленных последовательностей с оптимизацией разбиения на интервалы [5, 6]; сжатие изображений без потерь на основе кодирования длин серий [7]; предобработка изображений одномерными точечными отображениями [8-11], использование онтологического подхода для защиты данных при их пересылке и архивации [12], а также исследование технологий сжатия информации в рамках модели избыточности [13-16].

В указанном перечне следует выделить статьи [4, 5]. В работе [4] предложен метод, основанный на разделении цифрового изображения на разрядные двоичные изображения и представлении их двумерными марковскими цепями с двумя состояниями. Идея метода сжатия заключается в предсказании каждого элемента разрядных двоичных изображений и удалении из передаваемого потока правильно предсказанных пикселей. Восстановление непереданных пикселей осуществляется на приемной стороне с точностью 100%. В [5] предложено усовершенствование алгоритма VSEncoding сжатия без потери информации целочисленных данных, значения которых распределены преимущественно вблизи нуля. Такие данные получаются, например, при разностном кодировании целочисленных последовательностей, представляющих постепенно изменяющиеся величины (величины, принимающие близкие значения в соседних точках). По степени сжатия для этого вида данных предложенный в [5] алгоритм сжатия сравним с ZLib или превосходит его в режиме Z_BEST_COMPRESSION. Достигаемая степень сжатия составляет немногим более 22,31%. Он требует значительно меньше времени как при сжатии, так и при распаковке, поскольку характеризуется линейной вычислительной сложностью.

Исследования показали, что предпочтительным выходом из существующей проблемной ситуации является развитие математических методов представления информации. В частности, в настоящей статье предлагается метод сжатия многоразрядных данных на основе векторного их представления на основе использования псевдорегулярных чисел (ПРЧ). Метод доведен до программной реализации: демонстрационного прототипа комплекса, реализующего сжатие и декомпрессию многоразрядных данных.

Говоря о векторном представлении численных данных, уместно провести аналогию с векторной и растровой графикой [13-14]. Векторная графика, благодаря краткому математическому описанию графических объектов, требует заметно меньшей памяти ЭВМ, чем представление тех же объектов с помощью растровой графики. Недостатком векторной графики является невозможность представления большинства сложных растровых графических объектов, например, таких как фотографии. Использование векторного представления целых многоразрядных чисел, например, на основе разложения на множители или другого сжатого их аналитического представления с помощью аналитических формул, также дает возможности существенной экономии памяти, а также повышения безопасности хранения и передачи данных.

1. Описание метода представления данных

Состав исходных данных:

- двоичный код исходного файла произвольного объема и формата представления;
- носитель, предназначенный для хранения информации с объемом памяти не более 1 мегабайт.

Состав выходных данных: служебная информация и несжимаемый остаток исходного файла, представленный в виде вектора

<тип ПРЧ, длина ПРЧ, ….>,

записанного в файл объемом не более $20\log_2 n$, где n это количество разрядов сжимаемого двоичного кода, где ПРЧ – псевдорегулярное число (с псевдорегулярной двоичной структурой).

1.1. Теоретические положения и основные расчетные соотношения

Отправной точкой наших рассуждений служит двоично-десятичное преобразование Лейбница, предложенное в [17], согласно которому каждому натуральному числу соответствует его уникальная запись в двоичном виде:

$$N = \sum_{i=1}^{n} a_i 2^i . \qquad (1)$$

Для хранения и передачи информации К. Шеннон разработал метод сжатия, основанный на модели избыточности кода. Как правило, существующие методы и алгоритмы архивации (их насчитывается около 125) опираются на эту модель, которая позволяет обеспечить сжатие информации без потерь приблизительно в 30 раз. Этот предел сжатия был хорош во времена К. Шеннона, но обладает недостатками, отмеченными во введении.

В работах В.А. Котельникова, А.Н. Колмогорова [18] проведено исследование возможности математического вычисления произвольного натурального числа на основании некоей универсальной формулы и обоснована невозможность такого подхода для решения задач сжатия. Тем не менее, исследования в области теории чисел позволили сформулировать математический метод вычислений для реализации метода сжатия информации.

В настоящей статье рассматривается метод векторного представления целых чисел на основе разложения произвольного многоразрядного числа на сомножители. При этом, один из сомножителей является небольшим натуральным числом, другой – числом с псевдорегулярной структурой и разрядностью, совпадающей или близкой к разрядности сжимаемого числа. Для числа с псевдорегулярной структурой существует возможность короткой записи параметров для вычисления исходного числа.

1.2. Математическая модель

В полях, образованных натуральными числами, представленными двоичными кодами (от 0 до 2^n) существуют числа с псевдорегулярной структурой (табл. 1) – двоичные псевдорегулярные числа (ПРЧ) [19]. В этом поле они создают зоны (рис. 1), в которых они являются верхними и нижними границами. Кроме того, ПРЧ удобны для выполнения математических операций, так как для их запоминания требуется знать номер типа ПРЧ и длину ПРЧ (n) в битах. Внутренняя структура такого числа видна в столбце «вид» таблицы. Длина записи в сжатом виде для таких чисел $\sim 10 \cdot \log_2 n$. В таблице 1 и ниже ее показаны 15 типов ПРЧ, предложенных в [19].

Таблица 1 – Типы ПРЧ и формулы их генерации

№ типа ПРЧ	Вид ПРЧ (младший разряд справа)
1	10101010….
2	01010101….
3	01111111….
4	10000000….
5	…00001111…
6	…11110000…
7	11111111…
8	..00000001
9	...11111110

Дополнительные типы ПРЧ:

- 10 тип – 10 ….01 – это 10 + ПРЧ тип 4 (заполненный справа);
- 11 тип – 110…01 – это 11 + ПРЧ тип 4 (заполненный справа);
- 12 тип – 101111…1 – это 10 + ПРЧ тип 7 (заполненный справа);
- 13 тип – 01000…01 – это 01 + ПРЧ тип 4 (заполненный справа);

– 14 тип – 111…101 – это ПРЧ тип 7 (заполненный слева) +01 (т.е. отображение типа 12.
– 15 тип –000…000.

Пример для типа 14 показывает, что для записи служебной информации при сжатии заданного числа требуется указание тройки чисел:

<тип ПРЧ, его длина в битах,
направление заполнения (слева или справа)>.

При этом 14 тип ПРЧ фактически не нужен.

Исследования, проведенные на поле байта, показали, что все натуральные числа в действительности имеют не случайную двоичную внутреннюю структуру, а относятся к трем основным типам (табл. 2):

1. Числа с псевдорегулярной структурой (ПРЧ) (выделено зеленым цветом).
2. Числа, вычислимые из ПРЧ путем умножения на сомножитель m), при $m<<n$ (выделено белым цветом).
3. Числа, составные из ПРЧ (выделено коричневым цветом).

В двоичном поле байта (табл. 2):

– количество чисел с псевдорегулярной структурой – 70: 27,34375 %;
– количество чисел, вычислимых из ПРЧ – 150: 58,59375 %;
– количество чисел со структурой, составной из ПРЧ – 36: 14,0625 %.

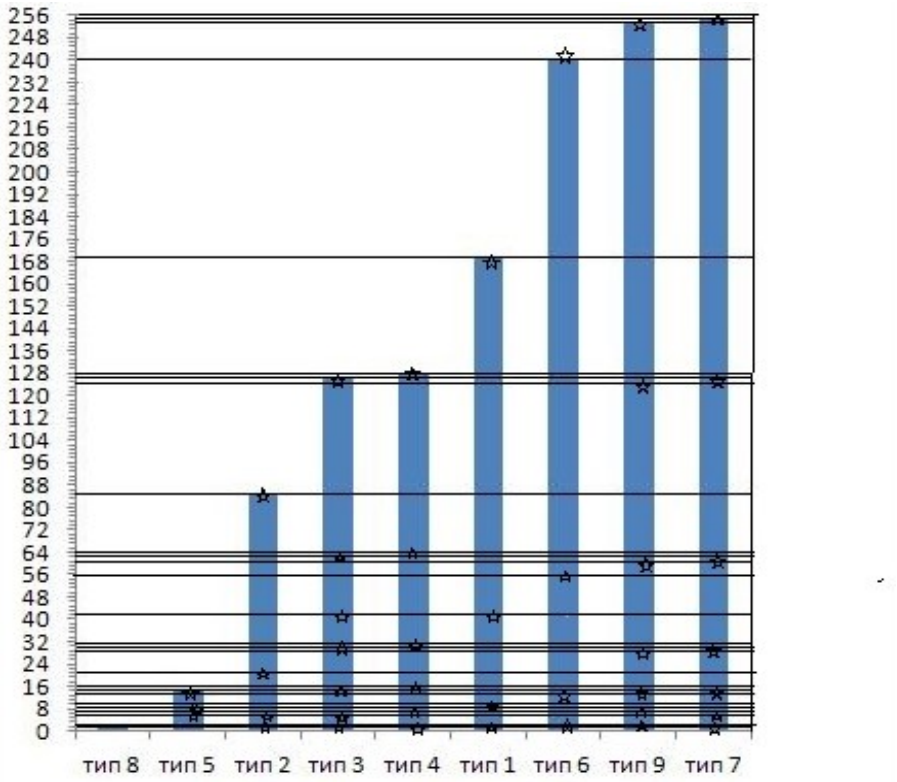

Рис. 1. Распределение ПРЧ в упорядоченном числовом поле на примере байта

Таблица 2 – Структура чисел в поле байта

1	17	33	49	65	81	97	113	129	145	161	177	193	209	225	241
2	18	34	50	66	82	98	114	130	146	162	178	194	210	226	242
3	19	35	51	67	83	99	115	131	147	163	179	195	211	227	243
4	20	36	52	68	84	100	116	132	148	164	180	196	212	228	244
5	21	37	53	69	85	101	117	133	149	165	181	197	213	229	245
6	22	38	54	70	86	102	118	134	150	166	182	198	214	230	246
7	23	39	55	71	87	103	119	135	151	167	183	199	215	231	247
8	24	40	56	72	88	104	120	136	152	168	184	200	216	232	248
9	25	41	57	73	89	105	121	137	153	169	185	201	217	233	249
10	26	42	58	74	90	106	122	138	154	170	186	202	218	234	250
11	27	43	59	75	91	107	123	139	155	171	187	203	219	235	251
12	28	44	60	76	92	108	124	140	156	172	188	204	220	236	252
13	29	45	61	77	93	109	125	141	157	173	189	205	221	237	253
14	30	46	62	78	94	110	126	142	158	174	190	206	222	238	254
15	31	47	63	79	95	111	127	143	159	175	191	207	223	239	255
16	32	48	64	80	96	112	128	144	160	176	192	208	224	240	256

2. Алгоритм представления информации на основе использования ПРЧ

Разложение числа в рассматриваемом методе можно представить графически нахождением позиции числа в двоичном поле, поделенном на m зон одинаковой размерности n, с помощью формулы вида:

$$Z = m \cdot n, \qquad (2)$$

где m – многоразрядное ПРЧ.

На основе вышесказанного может быть предложен ***алгоритм представления (сжатия)*** многоразрядных двоичных чисел с анализом их зависимости от псевдорегулярной структуры на ограниченном числовом поле и первых двух основных типов натуральных чисел, который включает следующие шаги:

1. Открытие исходного файла в бинарном виде (как файла с расширением *.bin).
2. Проверка двоичного числа, являющегося записью в файле, на псевдорегулярность путем сверки с 15 основными ***типами ПРЧ*** длины исходного файла n: если «да», то производится запись в служебный (выходной) файл что ***тип натурального числа*** 1 и записываются требуемые характеристики. Если «нет», то переходим к следующему шагу.
3. Производится деление двоичного числа на ПРЧ от 1 типа до 15 (см. табл. 1).
4. В каждом раунде цикла проверяется целочисленность результата деления: если «да», то производится запись в служебный (выходной) файл что ***тип натурального числа*** 2 и записываются требуемые характеристики. Если «нет», то переходят к следующему шагу.
5. Производится проверка на типы ПРЧ составного типа с 10 по 14.
6. Проверяется в каждом раунде цикла наличие 2 ПРЧ в составе двоичного кода: если «да», то производится запись в служебный (выходной)

файл что тип натурального числа 3 и записываются требуемые характеристики. Если «нет», то переходят к следующему шагу.

7. Если хоть один этап алгоритма приводит к записи в служебный выходной файл, по окончании операции END.

Формат записи сжатой информации в служебном файле приведен в табл. 3.

Таблица 3 – Формат служебной записи о результатах сжатия

Тип первого ПРЧ	Длина первого ПРЧ	Множитель m	Тип второго ПРЧ	Длина второго ПРЧ	Атрибуты файла

В служебный файл записывается информация о типе натурального числа, выраженного через параметры ПРЧ: тип, длина и начальный размер сжимаемого файла, его название и формат (расширение файла).

Обратное вычисление исходного числа (декомпрессия) заключается в создании пустого файла требуемого размера и формата, затем производится математическая операция вычисления двоичного числа на основании служебной информации о ПРЧ и запись его в сформированный файл.

Для примера отметим, что объем сжатой информации для 1 Тбайта определяется данными, приведенными в табл. 4.

Таблица 4 – Объем сжатой информации для 1 Тбайта и трех типов натуральных чисел

Номер типа натурального числа	Тип первого ПРЧ	Длина первого ПРЧ	Множитель m	Тип второго ПРЧ	Длина второго ПРЧ	Атрибуты файла	Итого
1	4 бита	64 бит	–	–	–	1024 бита	1092 бита
2	4 бита	64 бит	6 бит	–	–	1024 бита	1098 бит
3	4 бита	64 бит	–	4 бита	64 бит	1024 бита	1160 бит

Покажем, еще раз на простом примере, как это работает. Разложим исходное число (бинарного файла): $10485750_{10} = 10101010101010101010_2$ (или 699050_{10}) $\cdot 15_{10}$.

Для исходных чисел, содержащих очень большое количество знаков, выигрыш получается существенным.

3. Приведение произвольного двоичного числа к ПРЧ на основе двойного двоично-десятичного преобразования

Более того, проведенные исследования показали, что существуют и прямые соответствия «случайных» двоичных чисел и ПРЧ. То есть они могут быть *приведены* к псевдорегулярным числам путем двойного двоично-десятичного преобразования.

Пример:

1. Возьмем ПРЧ 4 типа: 1 0000 0000 0000 0000.
2. Представим его в десятичном виде: 65535.

3. Снова переведем в двоичный вид с записью каждого десятичного числа разным количеством бит:

110101101011101 (по 3 бита на число);
01100101010100110101 (по 4 бита на число);
0011000101001010001100101 (по 5 бит на число) и т.д.

Из приведенного примера видно, что число в п. 1 и различные числа в п. 2 и 3 есть одно и то же число, но первое число имеет псевдорегулярную структуру и поэтому отлично сжимается.

Заключение

Предложенный метод представления многоразрядных двоичных данных может найти применение для повышения устойчивости функционирования перспективных информационно-вычислительных систем. Актуальные вопросы и предложения по этой тематике рассматриваются в статьях [20-22].

Приведение произвольных двоичных чисел к псевдорегулярной структуре позволяет получать числа с нужными свойствами, а также обеспечивать структурную скрытность информации, например, при ее передаче в каналах связи. Физическая реализация ПРЧ кодов позволяет получать сигнал типа маяка, а также заменять сигнал, обозначающий «ноль» отсутствием сигнала, что особенно выгодно при большом количестве нулей в ПРЧ.

Новизна представленных в статье результатов состоит в том, что выявлено, что все натуральные числа не имеют случайную внутреннюю двоичную структуру, а могут быть вычислены или приведены к псевдорегулярному виду, который удобен для записи в сжатом виде. Векторное представление многоразрядных целых чисел дает возможность получения сжатой записи без использования модели избыточности. Таким образом, решена важная прикладная задача представления многоразрядных бинарных файлов с обеспечением сжатия, применяемого в криптографических примитивах, и повышения безопасности хранения и передачи данных.

Литература

1. Ветцель И. Эра «Больших данных» // Интерэкспо Гео-Сибирь. 2014. № 1. С. 15-19.

2. Rao K. R., Yip P. C. The Transform and Data Compression Handbook. – Boca Raton, CRC Press LLC, 2001. – 399 p.

3. Иваськив Ю. Л., Лещинский О. Л., Левченко В. В. Оценка качества обучающих множеств для нейронных сетей в задачах сжатия данных без потерь // Математические машины и системы. 2008. № 1. С. 91-96.

4. Петров Е. П., Харина Н. Л., Сухих П. Н. Метод сжатия многоразрядных спутниковых снимков без потерь // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 2. С. 203-210.

5. Хмельнов А. Е. Алгоритмы сжатия без потерь разностных целочисленных последовательностей при помощи оптимизации их разбиения на интервалы с постоянной битовой глубиной значений // Вычислительные технологии. 2015. № 3. С.75-98.

6. Keller Y., Averbuch A. Fast Motion Estimation Using Bidirectional gradient Methods [Электронный ресурс]. URL: http://www.eng.bui.ac.il/_kellery1/ publications /pdf/ optical_flow_ieee_final.pdf (дата обращения 01.06.2019).

7. Аль-Бахдили Х. К., Макейчик Е. Г., Цветков В. Ю. Сжатие полутоновых изображений без потерь на основе кодирования длин серий // Доклады Белорусского государственного университета информатики и радиоэлектроники. 2016. № 2. С.63-68.

8. Немировский В. Б., Стоянов А. К. Предобработка изображений одномерными точечными отображениями // Известия Томского политехнического университета. Инжиниринг георесурсов. 2011. № 5. С.107-111.

9. Акимов В. А. Дистанционные технологии в образовании. Алгоритмы сжатия информации и форматы данных для передачи текстовой, звуковой и видеоинформации // Известия Московского государственного технического университета МАМИ. 2013. Т. 2. № 4. С. 352-355.

10. Эльшафеи М. А., Сидякин И. М., Харитонов С. В. Исследование методов обратимого сжатия телеметрической информации // Вестник Московского государственного технического университета им. Н.Э. Баумана. Серия: Приборостроение. 2014. № 3. С. 92-104.

11. Муромский А. А., Тучкова Н. П. Использование онтологического подхода для защиты данных при их пересылке и архивации // Онтология проектирования. 2016. Т. 6. № 2. С.136-148.

12. Старобинец Д.Ю., Хомоненко А.Д., Гаврилова Н.А. Автоматический выбор параметров сжатия изображений с потерями на основе инвариантных моментов при дистанционном зондировании Земли // Современные проблемы дистанционного зондирования Земли из космоса. 2017. Т. 14. № 5. С. 26-36.

13. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ-МИФИ, 2003. – 384 с.

14. Хомоненко А. Д. Методы сжатия изображений: Учебное пособие. – СПб.: ПГУПС, 2009. – 31 с.

15. Сэломон Д. Сжатие данных, изображений и звука. – М.: Техносфера, 2005. – 368 с.

16. Григорьев А. Н., Шабаков Е. И., Дементьев А. Н., Романов А. А. Метод сокращения избыточности данных дистанционного зондирования из космоса // Известия высших учебных заведений. Приборостроение. 2016. Т. 59. № 1. С. 38-44.

17. Leibniz G., Explication de l'Arithmétique Binaire, Memoires de l'Academie Royale des Sciences. 1703. 85-92.

18. Колмогоров А. Н. Избранные труды в 6 томах. Том. 3. Теория информации и теория алгоритмов. – М.: Наука, 2005. – 263 с.

19. Воробьев Е. Г., Цехановский В. В. Псевдорегулярные числа в двоичных полях // Известия СПбГЭТУ «ЛЭТИ». 2014. № 2. С. 18-23.

20. Вихров Н. М., Ныркова А. П., Шнуренко А. А., Соколов С. С., Некрасова А. А., Полугина Ю. К. Современные методы оптимизации передачи

данных в информационно-вычислительных сетях на транспорте // Морской вестник. 2017. № 1 (61). С. 95-98.

21. Гарсиа Эскалона Х.А., Истомин Е. П., Колбина О. Н. Перспективы развития инфраструктуры пространственных данных с использованием современных технологий // Ученые записки Российского государственного гидрометеорологического университета. 2018. № 50. С. 130-136.

22. Воробьев Е. Г. Математические модели управления системой обеспечения доступности информации и оценки качества ее функционирования // Наукоемкие технологии в космических исследованиях Земли. 2019. Т. 11. № 2. С. 51-62.

References

1. Vetcel I. *Jera «Bol'shih dannyh»* [The Era of “Big Data”]. *Interjekspo Geo-Sibir*, 2014, no. 1, pp. 15-19 (in Russian).

2. Rao K. R., Yip P. C. *The transform and data compression handbook*. Boca Raton, CRC Press LLC, 2001. 399 p.

3. Ivaskov Yu. L., Leshchinsky O. L., Levchenko V. V. Ocenka kachestva obuchajushhih mnozhestv dlja nejronnyh setej v zadachah szhatija dannyh bez poter' [Evaluation of the Quality of Training Sets for Neural Networks in Lossless Data Compression Problems]. *Mathematical Machines and Systems*, 2008, no. 1, pp. 91-96 (in Ukrainian).

4. Petrov E. P., Kharina N. L., Chukaev K. N. Metod vydeleniya konturov ob”ektov na sputnikovykh snimkakh minimal’nymi vychislitel’nymi resursami [The Method of Allocation of Contours of Objects in Satellite Images Minimal Computing Resources]. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2016, vol. 13, no. 5, pp. 304-311 (in Russian).

5. Khmelnov A. E. Algoritmy szhatija bez poter' raznostnyh celochislennyh posledovatel'nostej pri pomoshhi optimizacii ih razbienija na intervaly s postojannoj bitovoj glubinoj znachenij [Lossless Compression Algorithms for Differential Integer Sequences by Optimizing their Division into Intervals with a Constant Bit Depth of Values]. *Computational Technologies*, 2015, no. 3, pp.75-98 (in Russian).

6. Keller Y., Averbuch A. Fast Motion Estimation Using Bidirectional gradient Methods. Available at: http://www.eng.bui.ac.il/_kellery1/publications/pdf/optical_flow_ieee_final.pdf (accessed 01 June 2019).

7. Al-Bahdili Kh. K., Makeychik E. G., Tsvetkov V. Yu. Szhatie polutonovyh izobrazhenij bez poter' na osnove kodirovanija dlin serij [Compression of Halftone Images Without loss Based on Length-Length Coding]. *Doklady BGUIR*, 2016, no. 2, pp. 63-68 (in Russian).

8. Nemirovsky V. B., Stoyanov A. K. Predobrabotka izobrazhenij odnomernymi tochechnymi otobrazhenijami [Preprocessing of Images by One-Dimensional Point Mappings]. *News of Tomsk Polytechnic University. Georesource engineering*, 2011, no. 5, pp. 107-111 (in Russian).

9. Akimov V. A. Distancionnye tehnologii v obrazovanii. Algoritmy szhatija informacii i formaty dannyh dlja peredachi tekstovoj, zvukovoj i videoinformacii

[Remote Technology in Education. Information Compression Algorithms and Data Formats for Transmitting Text, Audio and Video Information]. *Izvestiya MGTU "MAMI"*, 2013, no. 4, pp. 352-355 (in Russian).

10. Elshafei M. A., Sidyakin I. M. ,Kharitonov S. V. Issledovanie metodov obratimogo szhatija telemetricheskoj informacii [Research of Methods of Reversible Telemetric Information Compression]. *Herald of the Bauman Moscow State Technical University. Series Instrument Engineering*, 2014, no. 3, pp. 92-104 (in Russian).

11. Muromskii A. A., Tuchkova N. P. Ispol'zovanie ontologicheskogo podhoda dlja zashhity dannyh pri ih peresylke i arhivacii [Ontological Approach to the Data Protection for their Transfer and Archiving]. *Ontology of Designing*. 2016, vol. 6, no. 2 (20). pp. 136-148 (in Russian).

12. Starobinets D. Yu., Khomonenko A. D., Gavrilova N. A. Avtomaticheskij vybor parametrov szhatija izobrazhenij s poterjami na osnove invariantnyh momentov pri distancionnom zondirovanii Zemli [Automatic Selection of Compression Parameters for Lossy Images Based on Invariant Moments During Remote Sensing of the Earth]. *Current Problems in Remote Sensing of the Earth From Space*, 2017, vol. 14, no. 5, pp. 26–36 (in Russian).

13. Vatolin D., Ratushnyak A., Smirnov M., Yukin V. *Metody szhatija dannyh. Ustrojstvo arhivatorov, szhatie izobrazhenij i video* [Data Compression Methods. Device Archives, Image and Video Compression]. Moscow, DIALOG-MEPhI Publ., 2003. 384 p. (in Russian).

14. Khomonenko A. D. *Metody szhatija izobrazhenij* [Methods of image compression]. Saint-Petersburg, Petersburg State Transport University, 2009. 31 p. (in Russian).

15. Salomon D. *Data Compression: The Complete Reference*. Springer Science, 2004. 920 p.

16. Grigoriev A. N., Shabakov E. I. Dementiev A. N., Romanov A. A. Metod sokrashhenija izbytochnosti dannyh distancionnogo zondirovanija iz kosmosa [The Method of Reducing the Redundancy of Remote Sensing Data from Space]. *Proceedings of Higher Educational Institutions. Instrument Making*, 2016, vol. 59, no. 1, pp. 38-44 (in Russian).

17. Leibniz G. W. Explanation of Binary Arithmetic. *Die Mathematische Schriften*, ed. C. Gerhardt, Berlin 1879, vol.7, pp. 223-227 (in French).

18. Kolmogorov A. N. *Izbrannye trudy v 6 tomah. Tom. 3. Teorija informacii i teorija algoritmov.* [Selected Works in 6 volumes. Tom. 3. Information Theory and Theory of Algorithms]. Moscow, Nauka Publ., 2005. 263 p. (in Russian).

19. Vorobiev E. G., Tsekhanovsky V. V. Psevdoreguljarnye chisla v dvoichnyh poljah [Pseudo-regular numbers in binary fields]. *Izvestiya SPbGETU «LETI»*, 2014, no. 2, pp. 18-23 (in Russian).

20. Vikhrov N. M., Nyrkov A. P., Shnurenko A. A., Sokolov S. S., Nekrasova A. A., Polugina Yu. K. Sovremennye metody optimizacii peredachi dannyh v informacionno-vychislitel'nyh setjah na transporte [Modern Methods of Optimizing Data Transmission in Information And Computer Networks in Transport]. *Morskoy Vestnik*, 2017, no. 1 (61), pp. 95-98 (in Russian).

21. Garcia Escalona Kh. A., Istomin E. P., Kolbina O. N. Perspektivy razvitija infrastruktury prostranstvennyh dannyh s ispol'zovaniem sovremennyh tehnologij [Prospects for the Development of Spatial Data Infrastructure Using Modern Technologies]. *Proceedings of the Russian State Hydrometeorological University*, 2018, no. 50, pp. 130-136 (in Russian).

22. Vorobiev E. G. Mathematical Models of Ensuring Information Availability Management System and Quality Assessment of its Functioning. *H&ES Research*, 2019, vol. 11, no. 2, pp. 51-62 (in Russian).

Статья поступила 10 июня 2019 г.

Информация об авторах

Воробьев Евгений Германович – кандидат технических наук, доцент, заведующий кафедрой Информационная безопасность. Санкт-Петербургский государственный электротехнический университет им. В.И. Ульянова (Ленина). Область научных интересов: информационная безопасность, моделирование информационно-вычислительных систем. E–mail: vrbyug@mail.ru

Хомоненко Анатолий Дмитриевич – доктор технических наук, профессор, заведующий кафедрой Информационные и вычислительные системы. Петербургский государственный университет путей сообщения Императора Александра I. Область научных интересов: информационная безопасность, моделирование информационно-вычислительных систем. E–mail: khomon@mail.ru

Адрес: 190031, Россия, г. Санкт-Петербург, Московский пр., д. 9.

Vector Representation Models of Multi-Bit Binary Data Based on Pseudo-Regular Numbers

E. G. Vorobiev, A. D. Khomonenko

***Problem statement**: The most widely used lossless compression methods (code length encoding - Run Length Encoding, Huffman (normal and dynamic), arithmetic compression, dictionary compression methods), as a rule, use the statistical properties of individual bytes or bits of text or image. The most common lossless compression algorithms based on the use of variable length codes. In this case, compression achieved by assigning short codes to frequently occurring data elements, and long codes to rarely occurring elements. A significant limitation of this approach is a relatively small degree of data compression, for which no losses allowed. The compression operation is an elementary cryptographic operation. **The aim**: of the work is to develop mathematical operations for cryptographic primitives based on the vector representation model of multi-bit binary data using pseudo-regular numbers. **Novelty**: the proposed approach is to apply the properties of binary numbers associated with their mathematical and structural dependence on pseudo-regular numbers to get their short record. **Result:** is that this increases the degree of compression of large-sized binary data and increases the information security of their transmission and storage. In addition, a mathematical model and a description of the algorithm for converting multi-digit binary numbers to a pseudo-regular structure based on a binary binary-decimal transformation are proposed. The distribution of numbers with a pseudo-regular structure in ordered number fields shown. **Practical significance**: the presented solution implemented as a demonstration prototype of a software application that implements compression, storage and restoration (decoding) of large-sized binary data with analysis of dependence on pseudo-regular structure on a limited numerical field and three main types of natural numbers. The pro-*

posed solution can find practical application in order to back up data to increase information security and availability of advanced information and computing systems operating under conditions of influence.

***Keywords**: mathematical transformations for cryptographic primitives, vector representation of binary data, pseudo-regular numbers, data storage and recovery, information security.*

Authors Information

Evgeny Germanovich Vorobyev – Ph.D. of Engineering Sciences, Associate Professor, Head of the Department of Information Security. Saint Petersburg Electrotechnical University "LETI". Research interests: information security, modeling of information and computing systems. E-mail: vrbyug@mail.ru

Anatoly Dmitrievich Khomonenko – Dr. habil. of Engineering Sciences, Professor, Head of the Department of Information and Computing Systems. Emperor Alexander I Petersburg State Transport University. Area of scientific interests: information security, modeling of information and computing systems. E-mail: khomon@mail.ru

Address: Russia, 190031, Saint Petersburg, 9 Moskovsky pr.