

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ
INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS

УДК 004.627:504.064.36:519.688



DOI: 10.21822/2073-6185-2023-50-3-156-166 Оригинальная статья /Original article

Исследование и разработка методов сжатия данных для систем технического мониторинга

А. Г. Якунин

Алтайский государственный технический университет им. И.И. Ползунова,
656038, г. Барнаул, пр. Ленина, 46, Россия

Резюме. Цель. Целью исследования является развитие методов сжатия данных применительно к системам мониторинга. **Метод.** Исследования проводились преимущественно экспериментально, путем обработки данных, генерируемых информационно-измерительной системой Алтайского государственного технического университета в процессе проведения метеонаблюдений, температурного контроля и контроля расходования таких энергетических ресурсов как горячая и холодная вода, теплоснабжение. Для уплотнения хранимых данных предлагалось изменение их структуры и формы представления, степень сжатия находилась теоретически. **Результат.** Приводятся результаты исследований, проведённых с участием автора публикации, в области обработки, передачи и хранения данных. Рассматриваются методы потокового сжатия данных «на лету» по мере поступления как с потерями, так и без потерь, а также методы, основанные на оптимизации структуры баз данных, предназначенных для хранения собираемой в процессе мониторинга информации. Основное внимание уделено рассмотрению сжатия данных, создаваемых в процессе температурного мониторинга, поскольку именно контроль и наблюдение за температурными процессами наиболее распространён в различных технических системах управления технологическими процессами, системах учёта тепла как на стороне потребителя (ЖКХ), так и на стороне теплоснабжающей организации. **Вывод.** В подавляющем большинстве случаев сжатие без потерь позволяет сжать поступающую от системы мониторинга информацию более чем в 10 раз, а сжатие с потерями – существенно сильнее, причём без потерь прагматической ценности хранимой информации

Ключевые слова: сжатие данных с потерями, сжатие данных без потерь, потоковое сжатие, системы технического мониторинга, структуры данных, модификация RLE методов.

Для цитирования: А.Г.Якунин. Применение сжатия данных для систем технического мониторинга. Вестник Дагестанского государственного технического университета. Технические науки. 2023; 50(3): 156-166. DOI:10.21822/2073-6185-2023-50-3-156-166

Research and development of data compression methods for technical monitoring systems

A. G. Yakunin

I.I. Polzunova Altai State Technical University,
46 Lenin Ave., Barnaul 656038, Russia

Abstract. Objective. The aim of the study is to develop methods for data compression in relation to monitoring systems. **Method.** The studies were carried out mainly experimentally, by processing data generated by the information-measuring system of the Altai State Technical University in the process of meteorological observations, temperature control and control of the consumption of such energy resources as hot and cold water, heat supply. However, when it was proposed to change their structure and presentation form to compact the stored data, the degree of compression was found theoretically. **Result.** The results of studies conducted with the participation of the author of the publication in the field of data processing, transmission and storage are presented. Methods for streaming data compression “on the fly” as it arrives, both lossy and lossless, as well

as methods based on optimizing the structure of databases designed to store information collected during monitoring, are considered. The main attention is paid to the consideration of data compression created in the process of temperature monitoring, since it is the control and monitoring of temperature processes that is most common in various technical process control systems, heat metering systems both on the consumer side (housing and communal services) and on the side of the heat supply organization. **Conclusion.** In the vast majority of cases, lossless compression allows you to compress the information received from the monitoring system by more than 10 times, and lossy compression is much stronger, and without losing the pragmatic value of the stored information.

Keywords: lossy data compression, lossless data compression, streaming compression, technical monitoring systems, data structures, modification of RLE methods.

For citation: A. G. Yakunin. Research and development of data compression methods for technical monitoring systems. Herald of Daghestan State Technical University. Technical Science. 2023; 50(3): 156-166. DOI:10.21822/2073-6185-2023-50-3-156-166

Введение. Большинство поисковых систем и словарей определяют термин «технический мониторинг» как систематическое наблюдение за состоянием сооружения с целью оценки соответствия проектным решениям и нормативным требованиям. Применительно же к задачам наблюдения за параметрами технологических процессов часто применяют термин «технологический мониторинг». А поскольку наблюдение, например, за температурой не всегда связано с технологическими процессами, для таких наблюдений применяют термины «температурный мониторинг», «мониторинг температуры», а также «температурный контроль» или «контроль температуры» [1]. Поэтому в дальнейшем предполагается, что термин «технический мониторинг» включает в себя все другие термины, касающиеся систематического наблюдения и регистрации физических величин, описывающих изменение во времени состояния любых систем и протекающих в них процессов. Это могут быть технические и природные системы и объекты, включая здания и сооружения, а также присутствующие внутри или снаружи таких объектов среды: как жидкие, так твёрдые и газообразные.

Другим важным моментом является необходимость при проведении мониторинга фиксации на материальном носителе временной последовательности зарегистрированных в процессе мониторинга значений наблюдаемых параметров. В этом смысле следует различать понятия мониторинга и контроля. Как известно, при контроле тоже ведётся наблюдение состояния объекта, процесса или связанных с объектом или процессом сред. Однако цель контроля – проверка соответствия параметров, описывающих текущее состояние, заданным значениям или временным зависимостям и, при возникновении рассогласований, выработка управляющих воздействий, а также информирование обслуживающего систему персонала о возникновении в ней нештатных ситуаций. При мониторинге же фиксация результатов наблюдения выполняется с целью возможности восстановления таких наблюдений для их последующего обработки, подготовки и оформления каких-либо документов или для проведения исследований. Примером могут быть исследования, связанные с выявлением закономерностей изменения климата на земле или изменения солнечной активности, когда желательно иметь данные на протяжении как минимум десятилетий. Кроме того, такая фиксация может помочь при расследовании возникших в контролируемой системе аварийных ситуаций для выяснения причин их возникновения. Без неё нельзя представить и работу систем учёта отпуска тепловой или электрической энергии на тепловых и электрических станциях, когда по регламенту такая информация должна храниться не менее года. Аналогичные сроки предусмотрены и для журналов регистрации температурного режима холодильного оборудования.

Постановка задачи. Современные технические системы характеризуются большим числом контролируемых параметров, причём значение некоторых из них может достаточно быстро изменяться во времени, а зачастую ещё и может потребоваться передача большого потока измерительной информации между различными подсистемами систем управления

по телеметрическим каналам. В ряде случаев имеется необходимость фиксации результатов измерений за достаточно продолжительный временной интервал. Отсюда становится очевидной необходимость решения вопросов компактификации как передаваемых, таких и хранимых данных, генерируемых в таких системах. Сегодня вопросам сжатия данных посвящено огромное число работ и информационных ресурсов, например, [2-4], большинство из которых посвящены сжатию мультимедийной информации. В то же время стали появляться работы, посвящённые сжатию информации с сенсорных устройств в информационно-измерительных системах [5-7], а также применению для обработки специальных данных нетрадиционных алгоритмов, не применяемых для сжатия изображений и звука [8-10].

Целью данной работы является представление отдельных результатов исследований, связанных с разработкой эффективных высокопроизводительных методов и реализующих их алгоритмов сжатия информации.

Методы исследования. Для исследования использовались данные, полученные за период свыше 12 лет с информационно-измерительной системы университетского кампуса Алтайского государственного технического университета, подробно описанной в [11-13]. Изначально основным назначением данной системы был сбор данных о потреблении вузом энергоресурсов, однако впоследствии, после замены приборов учёта, акцент был перенесён на мониторинг метеоданных и температурных процессов, протекающих как внутри помещений вуза, так и в прилегающем к учебному корпусу воздушном пространстве. Именно данные температурного мониторинга и были использованы при проведении большинства экспериментов по сжатию данных. Вся собираемая системой информация о контролируемых в процессе мониторинга параметрах записывалась в базу данных, реализованную в системе управления базами данных MySQL. Структура базы представлена на рис. 1 и будет рассмотрена далее в статье.

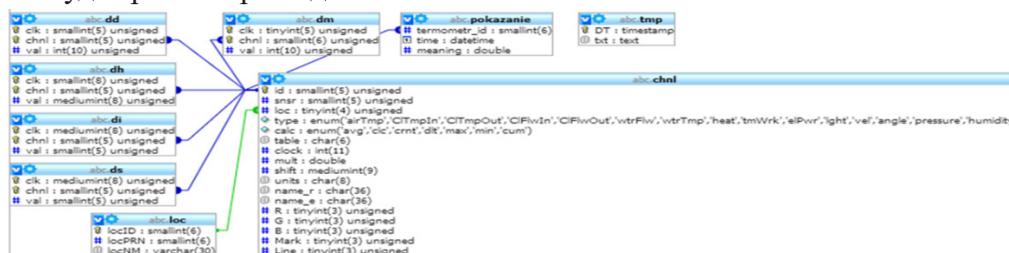


Рис. 1. Структура базы данных, использованной для хранения данных в описанной в [11-13] информационно-измерительной системе

Fig. 1. The structure of the database used to store data in the information-measuring system described in [11,13]

Хранимые в базе данные затем использовались для проведения экспериментов по их сжатию как известными методами, а также новыми методами, большинство из которых являлись модификацию уже существующих алгоритмов. При этом оценивалась эффективность сжатия разработанного метода для различного характера изменения во времени наблюдаемых процессов относительно классических алгоритмов сжатия, в том числе и относительно различных архиваторов, используемых в вычислительных устройствах.

Обсуждение результатов. Методы на основе нормировки и структуризации данных. При решении задач технического мониторинга оптимизировать объем хранимых данных можно за счёт применения их нормализации и структуризации, учитывающей вид измеряемой величины и динамику её изменения во времени. Частично эта задача была решена в реализованной в системе мониторинга [11-13] и показанной на рис. 1 базе данных. Главной таблицей базы является таблица “chnl”, хранящая информацию обо всех измерительных каналах системы. Применительно к теме данной работы в ней задействованы поля “id”, “calc”, “table”, “clock”, “mult” и “shift”. Поле “id” задаёт номер канала, с которого поступают данные, а поле “table” указывает на одну из предназначенных для хранения поступающих из канала исходных или обработанных данных таблиц “ds”, “di”, “dh”, “d”, “dd”, “dm” (соответственно, с кратным секундным, минутным, часовым, суточным или месячным

интервалом). В этих таблицах поле “chnl” соответствует номеру канала, указанному в поле “id” таблицы “chnl”, а в полях “val” и “clc” хранятся нормализованные значения исходных или обработанных результатов измерений и номера отсчётов n_i .

В системе минимальным неделимым временным интервалом является интервал базового такта $\Delta_t=30c$. Как показали исследования, в более частой временной дискретизации для регистрируемых в системе физических величин нет необходимости. Все остальные временные интервалы, используемые для дискретизации измерений, проводимых в разных каналах, кратны значению $30c$ и коэффициент кратности k_p , задающий число базовых тактов между смежными отсчётами контролируемого параметра, хранится в поле “clock”. Поэтому в таблицах “ds”, “di”, “dh”, “dd”, “dm” реальное время t_i , в которое был сделан отсчёт, находится из выражения $t_i=t_0+k_i \cdot n_i \cdot \Delta_t$, где t_0 – дата и время инициализации базы данных, заданное в файле инициализации. Такое решение позволило существенно сократить место, отводимое в базе под хранение информации о времени получения отсчётов [14]. Так, если бы поле “clc” хранило фактическое время проведения соответствующего измерения, то его размер в СУБД MySQL, его размер был бы равен 8 байтам, тогда как, как видно из табл.1, при предложенном варианте хранения временных отсчётов для этого более чем достаточно длины поля в 2-3 байта.

Таблица 1. Зависимость максимально возможного времени хранения (заданное в годах), от интервала выборки данных и заданного в байтах размера поля “clk”

Table 1. Dependence of the maximum possible storage time (specified in years), on the data sampling interval and the size of the “clk” field specified in bytes

Кол-во байт/ Number of bytes	Интервал выборки данных/ Data sampling interval									
	1c	10c	30c	1мин	5мин	10мин	30мин	1 час	1 сутки	1 месяц
1									0.7	21
2						1.3	4	7.5	180	
3	0.53	5.3	10.6	32	160	133				
4	136									

Стоит отметить, что при таком подходе к выбору длины поля нужно подходить с большой осторожностью. В частности, в описанной в [11-13] информационной системе на момент ее инициализации в 2009 году изначально были выбраны размеры полей, выделенные в таблице серым цветом. В результате в части таблиц, а именно, в таблицах “ds” и “di” по прошествии соответственно 5 и 10 лет возникла необходимость в увеличении размера поля на 1 байт (на рис.1 для таблиц показаны изначально выбранные типы полей). Более того, в процессе работы с базой выяснилось, что значение t_0 лучше вносить не в отдельную переменную, а непосредственно указывать её в соответствующем поле соответствующего канала в таблице “chnl”, поскольку в процессе работы часть каналов были закрыты, а часть, наоборот, введены в действие в более поздний срок.

Поле “calc” в таблице “chnl” указывает на характер обработки поступающей из канала измерительной информации, результаты которой будут храниться в системе. Такими операциями могут быть усреднение данных, выборка максимального, минимального или накопленного значения, либо же отказа от выполнения операций и сохранение данных в исходном виде (когда значение поля “calc” равно “crnt”).

В зависимости от вида обработки данных и интервала их выборки после выполнения такой обработки они записываются в поле “val” в одну из таблиц “ds”, “di”, “dh”, “dd”, “dm” в соответствии со значением поля “table” таблицы “chnl”. В данной системе в таблицу “ds” записывались результаты температурных измерений и скорости ветра, в таблицу “di” – сведения о давлении и влажности, в таблицу “dh” – показания теплосчётчиков и водосчётчиков, а в таблицах “dd”, “dm” – накопленные показания приборов учёта соответственно за сутки и за месяц. В любом случае, обязательной операцией перед записью является операция нормализации данных, так как они поступают в систему из измерительных каналов в виде вещественных чисел, соответствующих действительному значению измеряемого параметра в соответствии в заданной для него в поле “units” единице измерения. Эти числа пересчитываются в целые нормализованные значения по формуле $y_i=round(k_x \cdot x_i + y_0)$, где x_i

– поступающие из канала данные, k_x – мультипликативный коэффициент масштабирования исходного значения параметра, y_0 – аддитивный коэффициент его смещения, а **round** – функция округления результата вычислений до целого значения. Значения подбираются так, чтобы самому минимальному значению измеряемого параметра соответствовало нулевое значение y_p , а максимальному – значение, близкое к 2^{8n} , где n – число байт, отводимых под хранение результатов измерения наблюдаемого параметра. В результате такой нормализации для хранения данных потребуется поле длиной не 8, а всего 2-3 байта, поскольку крайне редко погрешность измерения бывает менее 0.1-0.001%, то есть при минимуме вычислительных операций можно сжать данные в 2-4 раза.

Другой особенностью предложенной структуры является возможность хранения одного и того же параметра сразу в нескольких таблицах., в зависимости от характера его обработки. Так, в таблице “ds” можно хранить исходные текущие показания водосчетчика, в таблице “dh” – накопленные значения расхода за 1 час, в таблице “dd” – накопленные суточные значения расхода и в таблице “dm” – месячные расходы. Конечно, объем хранимых при этом данных хотя и незначительно, но возрастает, однако, при этом существенно повышается оперативность получения необходимых сведений. В дальнейшем предполагается дальнейшее совершенствование структуры базы данных, целью которого являлось сочетание высокой степени сжатия и быстрого доступа к данным. Для этого предполагается применять рассмотренные далее методы для максимального сжатия архивных данных прошлых периодов (более года) и использовании менее ресурсоёмких методов сжатия для доступа к оперативным данным текущих периодов. Такой подход вполне оправдан, так как данные прошлых лет менее востребованы и чаще используются для проведения исследовательских работ, когда время доступа к данным не имеет принципиального значения, поскольку оно будет всё равно много меньше времени, необходимого для выполнения самого исследования.

Методы сжатия с применением разностных схем и RLE — алгоритмов. Дальнейшего повышения степени сжатия можно добиться за счёт применения разностных схем [15]. Принцип такого сжатия поясняет рис.2.

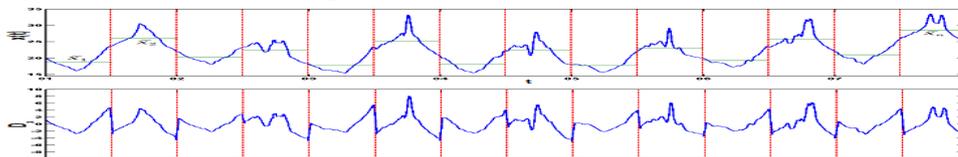


Рис. 2. Временные диаграммы циклического изменения суточной уличной температуры $x(t)$ и разностные сигналы $D(t)$ между этими изменениями и её среднесуточными значениями X_i

Fig. 2. Time diagrams of cyclic change in daily street temperature $x(t)$ and difference signals $D(t)$ between these changes and its average daily values X_i

Из него видно, что вместо исходного сигнала хранятся его отклонения от среднего значения, взятые на некотором временном интервале (в данном случае, сутки). Поскольку отклонения от среднего, как видно из рис., в 2–4 раза меньше значений исходного сигнала, это позволяет сократить размерность необходимой для хранения разрядной сетки на 1–2 двоичных разряда. Можно уменьшить величину отклонений ещё существеннее, если сократить интервал усреднения. Однако при этом не следует забывать, что помимо отклонений необходимо хранить и сами средние значения, что снижает степень сжатия.

Кроме того, если в результатах наблюдений имеется тренд или значительные колебания, как это имеет место на рис.2, разностные значения быстро становятся слишком велики и будут мало чем отличаться от начального опорного значения. Именно критерий достижения разностных значений заданного порога и будет являться главной причиной ограничения числа членов временного ряда. Поэтому базовый разностный метод малоэффективен и на практике чаще применяются схемы, когда запоминаются не отклонения от одной опорной точки (будь то среднее значение или какой-то абсолютный отсчёт), а разность значений

смежных отсчётов. На рис. 3 в качестве примера представлены различные типовые графики изменения температуры на улице и в различных точках помещения. Измерение температуры проводилось с использованием датчиков DS18B20 с температурным разрешением 0.0625°C .

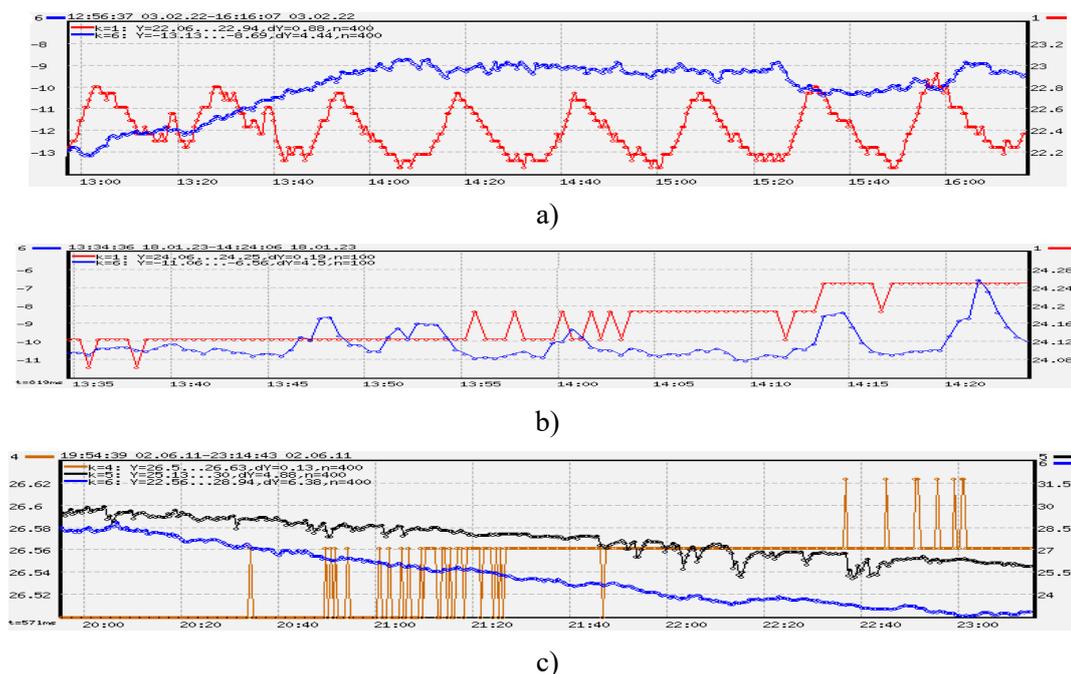


Рис 3. Диаграммы временного изменения уличной температуры (синий график, канал 6), температуры в различных точках комнаты (красная линия, канал 1 – температура у потолка, коричневая линия, канал 4 – у пола) и в чердачном помещении (канал 5, черная линия). а) – работает электрообогреватель, создающий в помещении регулярные температурные осцилляции, б) сильные порывы ветра, приводящие из-за наличия в воздушных массах микро-турбулентностей к температурным скачкам до $2^{\circ}\text{C}/\text{с}$, в) – штиль, плавное похолодание к ночи

Fig. 3. Diagrams of time changes in outdoor temperature (blue graph, channel 6), temperature at various points in the room (red line, channel 1 - temperature at the ceiling and brown line, channel 4 - at the floor) and in the attic (channel 5, black line). a) - an electric heater is working, creating regular temperature oscillations in the room, b) strong gusts of wind, leading to temperature jumps up to $2^{\circ}\text{C}/\text{s}$ due to the presence of micro-turbulences in the air masses, c) - calm, gradual cooling by night

Как видно из рис.3, даже при самых резких температурных колебаниях перепад значений смежных отсчётов не превышает 32 уровней квантования (рис.3б, последний правый выброс температуры). Поэтому под один отсчёт достаточно отводить не 10, а только 5 бит, что позволяет почти вдвое сократить объем хранимых данных практически без снижения скорости их последующей выборки.

Для повышения надёжности и сокращения времени, затрачиваемого на восстановление данных, последовательность разностных значений должна периодически прерываться и начинаться заново, чтобы образующий серию отсчётов числовой ряд имел ограниченную длину. Тогда потеря или ошибка единственного начального абсолютного значения не приведёт к полной потере информации обо всех входящих в наблюдение контролируемой физической величины сериях отсчётов, и для восстановления значений отсчётов за последний период наблюдений не потребуется выполнять полный пересчёт, начиная с самого первого значения временного ряда.

Структура базы данных, предназначенная для хранения зарегистрированных и упакованных таким образом отсчётов, приведена в табл.2.

Образуемый нормализованными значениями измеряемой величины временной ряд разбивается на отдельные серии и в табл. “Reference” для каждой серии фиксируется время измерения и абсолютное значение первого входящего в серию отсчёта, а также общее количество таких отсчётов и значение индекса первого входящего в серию разностного значения. Разностные значения и их индексы хранятся в табл. “Data”.

Таблица 2. Структура данных для хранения серий разностных отсчётов
Table. 2. Database structure for storing series of difference readings

Таблица “Data”

Ключ Key	Имя поля Field Name	Описание Description
PK	ID	Номер элемента
	Offsets	Разностные значения

Таблица “Reference”

Ключ Key	Имя поля Field Name	Описание Description
PK	Start_time	Начало серии
FK	Series_ID	Номер первого элемента серии (внешний ключ)
	Init_Val	Initial Value – абсолютное значение первого элемента в серии разностных значений
	Ser_Len	Series length - число элементов, входящий в одну серию данных

При этом время начала серии для компактификации данных тоже лучше записывать не UTCформате, а в виде числа тактов, прошедших с момента активации канала. Для хранения данных можно обойтись и одной таблицей “Reference”, если поле “Series_ID” заменить бинарным полем переменной длины, в которое и будет записываться вся серия разностных значений. При этом длина серии будет определяться лимитом времени, необходимым для восстановления исходного временного ряда. Другой причиной ограничения длины серии может явиться появление резкого изменения значения измеряемой величины, когда для записи соответствующей ему разности окажется недостаточно числа бит, выделяемых под хранение разностного значения, число которых находится путём проведения статистических исследований для конкретного регистрируемого процесса. Очевидно, такое очередное значение окажется первым элементом очередной серии регистрируемых данных.

С другой стороны, как видно из рис.3, эти значения могут иметь не только резкие выбросы, но и на протяжении достаточно долгого времени оставаться постоянными (рис.3б, канал 1), либо постоянными сохраняются разности смежных отсчётов (рис.3а, канал 1, отдельные участки на склонах сигналов треугольной формы). Отсюда напрашивается вывод о целесообразности дополнения метода сжатия путём хранения разностных значений методом, известным в литературе как алгоритм длин серий или алгоритм RLE (run-length encoding) [3]. Суть его состоит в том, что входящие в зарегистрированную последовательность результатов измерений участки, состоящие из следующих подряд неизменных значений, заменяются парой чисел. Одно из них задаёт число таких значений, а другое равно самому этому значению. Поскольку при мониторинге даже одного и того же процесса характер его изменения во времени может быть совершенно различным, было предложено модифицировать стандартный алгоритм путём введения в последовательность пар «значение – число повторов» дополнительных признаков, позволяющих в этой последовательности менять число бит, отводимых под запись максимально возможного числа повторов, и разностного значения, а также под общее число байт, отводимых под запись следующих рядом таких пар.

Например, если разностные значения лежат в пределах восьми уровней квантования с учётом знака (от -3 до +4), а число повторов одинаковых значений не превышает 16, то для записи одной пары «значение – число повторов» потребуется $3+4=7$ бит. Тогда в 3 байта можно записать 3 таких пары и 3 бита останется под запись кода признака, задающего характер компоновки данных в следующие байты. При этом, независимо от способа компоновки информации в байты, старший знаковый бит старшего байта должен выполнять функцию флага, указывающего на то, что следующие за ним биты (от одного до трёх) будут задавать характер компоновки и число следующих за ним информационных байт. Такой гибридный подход позволил за счёт непрерывного анализа поступающих данных выполнять смену формата упаковки результатов наблюдения с учётом динамики их изменения и, тем самым, обеспечить степень сжатия, сопоставимую, а в ряде случаев (в зависимости от характера регистрируемого процесса) и превосходящую степень сжатия современных архиваторов.

Таблица 3. Результаты сжатия данных температурных каналов системы мониторинга обычными архиваторами и предложенным алгоритмом (Org)*
Table 3. The results of data compression of the temperature channels of the monitoring system by conventional archivers and the proposed algorithm (Org)*

Bzip2	LZMa	LZMa2	PPMd	GZIP	XZ	RAR	ZIP	Deflate	Org
Резкие перепады уличной температуры/ Sudden changes in outside temperature									
1406	1597	1602	1723	1725	1542	1403	2137	1806	1820
Плавные изменения комнатной температуры/ Smooth changes in room temperature									
992	1138	1153	1220	1240	1091	1052	1561	1294	980

*Исходные данные – 14000байт нормализованных 7000 отсчётов. Выделены наиболее эффективные алгоритмы сжатия/Initial data - 14000 bytes of normalized 7000 readings. The most efficient compression algorithms are highlighted

Проведённые исследования, отдельные результаты которых представлены в табл. 3 подтвердили, что без потери информации можно достичь степень сжатия уже нормализованных данных в 14.2 раза при благоприятных условиях наблюдений и в 7.7 раза – при неблагоприятных [6]. При этом, в отличие от классических алгоритмов, метод обеспечивает потоковое, а не блочное сжатие, алгоритмически прост и не требует для реализации значительных вычислительных ресурсов.

Методы сжатия с потерями. Как известно, любые измерения имеют погрешности, поэтому сохранять информацию с высокой степенью точности нецелесообразно, если погрешность измерения превышает погрешность его округления при записи результата. Не имеет смысла сохранять результат измерений с высокой точностью и тогда, когда из-за наличия внешних шумов, помех и различных других дестабилизирующих факторов в нём присутствуют ошибки и отклонения, превышающие приборную погрешность. Особенно это наглядно видно при представлении результатов измерений в цифровом виде. Так, на рис.3b, канал 1 и на рис 3с., канал 4 на протяжении достаточно долгого времени измеренное значение остаётся в целом неизменным или изредка переходит с одного уровня квантования на другой, но при этом непрерывно флуктуирует в пределах одного уровня квантования. В результате, использование ранее рассмотренных методов сжатия на базе RLE алгоритмов становится малоэффективным, так как протяжённость участков с неизменными значениями из-за таких флуктуаций становится очень непродолжительной. Избавиться от таких флуктуаций можно путём усреднения значений временного ряда по нескольким точкам, например, методом скользящего среднего, либо путём аппроксимации участков фиксированной длины полиномами или тригонометрическими функциями, либо за счёт применения нелинейной фильтрации [17], суть которой поясняет рис. 4.

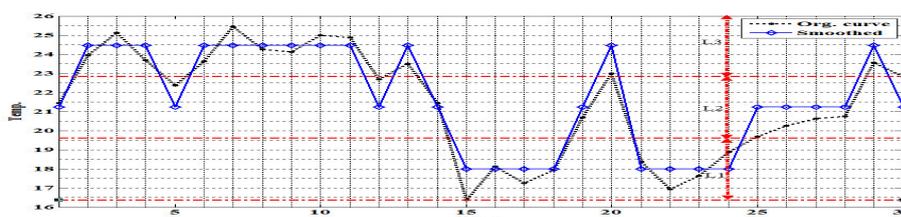


Рис. 4. Применение нелинейной фильтрации для сокращения числа изменений регистрируемой физической величины

Fig. 4. Application of non-linear filtering to reduce the number of changes in the registered physical quantity

Для реализации такой фильтрации весь диапазон изменения физической величины заменяется на ряд слоёв толщиной L_p , в пределах которых любые фактические значения измеряемой величины заменяются одним значением, соответствующим середине слоя. При этом, для усиления эффекта сглаживания, можно потребовать, чтобы переход сигнала из одного слоя в другой выполнялся лишь после того, как в этом слое окажется достаточное число очередных значений сигнала. Ещё лучший эффект сглаживания с минимальным искажением исходного сигнала даёт введение между слоями области неопределённости. При нахождении сигнала в этих областях считается, что его значение остаётся прежним, то есть в том слое, в котором он находился до перехода в область неопределённости.

Суть данного подхода поясняет рис.5, где предполагается, что для подтверждения перехода в другой слой достаточно нахождения в нём трех следующих подряд отсчётов.

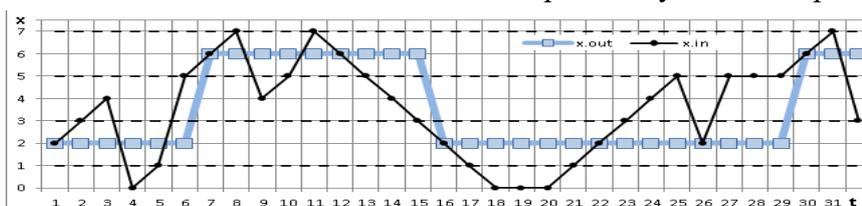


Рис. 5. Повышение эффективности нелинейной фильтрации за счёт введения областей неопределённости ($x=3\dots5$) и накопления трёхкратного подтверждения о необходимости перехода в другую область ($t=6\dots8$; $15\dots17$ и $29\dots31$)

Fig. 5. Increasing the efficiency of nonlinear filtering by introducing areas of uncertainty ($x=3\dots5$) and accumulating triple confirmations about the need to move to another area ($t=6\dots8$; $15\dots17$ and $29\dots31$)

Проведённые исследования показали, что переход к описанным методам сжатия позволяет более чем в 60 раз сократить объем записываемой информации [18].

Вывод. Рассмотренные в работе новые и модификации известных методов и подходов к сжатию измерительной информации показали, что их применение как в части организации хранения данных, так и в части степени сжатия информации, скорости её сжатия и извлечения превосходят известные методы, описанные, например, в работах [19-23]. Данный эффект достигается за счёт того, что перед этапом применения классического RLE – алгоритма и алгоритма записи разностных значений данные предварительно анализируются с целью выбора формата хранения данных, то есть упаковки данных о нескольких отсчётах в целое число байт. При этом дополнительно в поток данных инкапсулируются флаги – признаки, определяющие формат хранения данных.

Однако, наибольший эффект достигается за счёт применения методов сжатия с потерями, когда перед вышеописанным методом сжатия без потерь вводится этап предварительной нелинейной фильтрации, направленной на минимизацию флуктуаций сигнала в пределах младших разрядов цифрового представления результатов наблюдений, за счёт чего прагматическая ценность информации практически полностью сохраняется, но степень сжатия возрастает с 7-10 до 50–60. Ещё одним перспективным направлением является применение для сжатия результатов мониторинга методов, основанных на применении модели E-слоя [24,25] и, в частности, для сжатия данных «на лету» при передаче измерительной информации, как это предлагается в работе [26].

Библиографический список:

1. Контроль температурного режима, контроль температуры//<https://ruglonass.ru/kontrol-temperaturnogo-rezhimana-skladakh.php> (01.06.2023)
2. Всё о сжатии данных, изображений и видео// <https://www.compression.ru/> (01.06.2023)
3. Solomon, D. Data Compression: The Complete Reference: Fourth edition/ David Solomon.- London: Springer, 2007.- 1093p.
4. Хахина А.М. Современные алгоритмы архивации/А.М. Хахина , Н.Ю. Константинов // Заметки ученого - 2020г. №2. - С.189 -192. - <https://e.lanbook.com/reader/journalArticle/704000>
5. Лихтциндер, Б.Я. Адаптивное сжатие измерительных данных в беспроводных сенсорных сетях/Б.Я. Лихтциндер, Ю.О. Бакай // Измерение. Мониторинг. Управление. Контроль - 2021г. №1. - С.52-57 <https://e.lanbook.com/reader/journalArticle/603370>
6. Богачев, И.В. Геометрический подход к сжатию данных телеметрических систем/ И.В.Богачев, А.В. Левенец , Ун Чье Ен// Информатика и системы управления - 2015г. - С.16 -22. - №4 <https://e.lanbook.com/reader/journalArticle/346090>
7. Артюшкин А.Б., М.А. Куксенко, А.П. Пантенков. Экономное кодирование как метод повышения скорости передачи информации в телеметрических системах//Вестник Российского нового университета. Серия Сложные системы модели, анализ и управление.2020г. №1. С.43 -54. - <https://e.lanbook.com/reader/journalArticle/782456>
8. Бакулина, М.П. Эффективное кодирование длин серий при факсимильной передаче данных по сети//Ползуновский вестник 2013г. №2. С.257 -259. <https://e.lanbook.com/reader/journalArticle/405229>
9. Минкин, А. С. Сжатие гиперспектральных данных методом главных компонент/А. С. Минкин, О. В. Николаева, А.А. Руссков// Компьютерная оптика. 2021г. №2. С.235-246. - <https://e.lanbook.com/reader/journalArticle/595721>
10. Ярмоленко, А.С. Применение теории вейвлетов при сжатии и фильтрации геоинформации/А.С. Ярмоленко, О.В. Скобенко//Записки Горного института - 2018г. - С.612 -623. - №234 <https://e.lanbook.com/reader/>

journalArticle/71014

11. Сучкова, Л.И. Масштабируемые программно-технические решения для экологического и технического мониторинга/Сучкова Л.И., Х.М.Хуссейн, М.А. Якунин, А.Г. Якунин// Ползуновский Альманах - 2013г. №1. - С. 75- 82. <https://e.lanbook.com/reader/ journalArticle / 379151>
12. Hussein H.M. Design and implementation of weather and technology process monitoring systems/H.M. Hussein, R.V.Kuntz, L.I. Suchkova, A.G.Yakunin//Известия Алтайского государственного университета, 2013, No. 1-1(77), С. 210-214.
13. Yakunin, A.G. Hardware-software and algorithmic provision of multipoint systems for long-term monitoring of dynamic processes/A.G. Yakunin, H.M. Hussein//IOP Publishing IOP Conf. Series: Journal of Physics: Conf. Series 881 (2017) DOI: 90.1088/1742-6596/881/1/012028
14. Hussein, H.M. Ways for improving methods of data storing in monitoring systems / H.Hussein, L.I. Suchkova, M.A. Yakunin // Ползуновский альманах, № 2.–Барнаул: Изд-во АлтГТУ, 2012.– С. 48-50.
15. Hussein, H. M. Data Differencing Method To Optimize Data Storing In Weather Monitoring System / H.M. Hussein A.G. Yakunin // Ползуновский вестник, 2013.- № 2. – С.65-68. URL: https://journal.altstu.ru/media/f/old2/prv2013_02/pdf/065hussein.pdf
16. Мельников Е.М. Исследование эффективности методов сжатия без потерь данных температурного мониторинга / Е.М. Мельников., А.Г. Якунин // Материалы XX Международной научно-технической конференции «Измерение, контроль, информатизация» - 2019. - Барнаул, АлтГТУ, 2019 г. - С.32-38 <https://elibrary.ru/item.asp?id=41166261>
17. Hussein, H.M. Simple Curve Smoothing Methods For Weather Monitoring System/ H. Hussein, A. Yakunin // Materiali XI Mezinarodni vedecko-practicka conference «Veda a Technologie: krok do budoucnosti-2015», Dil 17. Technicke vedy. - Praha, 2015. - pp. 73-76.
18. Hussein, H. A Comparison of Data Compression Methods for Solving Problems of Temperature Monitoring/H.Hussein , A. Yakunin, L Suchkova// MATEC Web of Conferences, 2016, V. 79(2016). URL: <http://dx.doi.org/ 10.1051/mateconf/20167901076>.
19. Zhen, Ch. Design and Realization of Data Compression in Real-Time Data-base/Ch. Zhen and B. Ren. *IEEE International Conference on Computational Intelligence and Software Engineering, China*, 11 – 13 Dec. 2009; 340-243.
20. Ray G. Data Compression in Databases//Master’s Thesis, Dept. of Computer Science and Automation, Indian Institute of Science, June 1995.
21. Yang, H. On the performance of data compression algorithms based upon string matching/H. Yang// IEEE Transactions on Information Theory, Jan 1998, Vol. 44, Issue 1, pp.47- 65.
22. Lin, M. A New Architecture of a Two-Stage Lossless Data Compression and Decompression Algorithm/M. Lin and Y. Yi Chang//IEEE transactions on VLSI, 2009; 17(19): 1297- 1303.
23. Tamrakar, A A Compression Algorithm for Optimization of Storage Consumption of Non Oracle Database/Abha Tamrakar and V. Vinti Nanda//Research Gate, July 2012, Vol. 1, Issue 5, pp. 39-43.
24. Yakunin, A. Application of the E – layer model for solving the problems of parametric estimation in measuring devices/A. Yakunin , L. Suchkova// MATEC Web of Conferences 155, 01020 (2018) DOI: <https://doi.org/10.1051/mateconf/201815501020>
25. Suchkova, L.I. Application of E-Regions of the Parameters of the Model Function of the Signal for the Extraordinary Situations Detection/L.I. Suchkova, A.G. Yakunin//in “2018 International Multi-Conference on Industrial Engineering and Modern Technologies” (FarEastCon, 3-4 Oct. 2018, Vladivostok, Russia, added to to IEEE Xplore: 07 January 2019), pp.1-6, URL: <https://ieeexplore.ieee.org/document/8602766>, DOI: 10.1109/FarEastCon.2018.8602766
26. Сулова С.В. Способ выделения тренда нестационарного процесса с адаптацией интервалов аппроксимации/Сулова С.В., Якунин А.Г.//Патент России № 2645273. 2018. Бюл. № 5.

References

1. Temperature control, control of temperature. <https://ruglonass.ru/kontrol- temperaturnogo-rezhima-na-skladakh.php> (01.06.2023) (In Russ)
2. All about data, image and video compression. <https://www.compression.ru/> (01.06.2023) (In Russ)
3. Solomon D. Data Compression: The Complete Reference: Fourth edition/ David Solomon.- London: Springer, 2007;1093.
4. Hahina A.M.: Modern archiving algorithms. Notes of a scientist 2, 189 -192 (2020). - <https://e.lanbook.com/reader/ journalArticle/704000> (In Russ)
5. Lihtsinder B.Ya, Bakai Yu.O.: Adaptive measurement data compression in wireless sensor networks. Measurement. Monitoring. Control. Cheking 1, 52-57 (2021). <https://e.lanbook.com /reader/journalArticle/603370> (In Russ)
6. Bogachev I.V., Levenets A.V. Chie Yen Un: Geometric approach to data compression of telemetry systems. Informatics and control systems 4, 16-22 (2015).<https://e.lanbook.com/ reader/ journalArticle/346090> (In Russ)
7. Artyushkin A.B. Kuksenko M.A., Pantenkov A.P. Economical coding as a method of increasing the speed of information transmission in telemetry systems. Bulletin of the Russian New University. Series Complex Systems Models, Analysis and Control 2020;1:43-54 <https://e.lanbook.com/reader JournalArticle/782456>

8. Bakulina M.P. Efficient coding of run length encoding during facsimile data transmission over the network. Polzunovskiy vestnik 2013;2,257-259 <https://e.lanbook.com/reader/journalArticle/405229> (In Russ)
9. Minkin A.S., Nikolaeva O.V., Russkov A.A.: Compression of Hyperspectral Data by Principal Component Method. Computer optics 2021; 235-246 <https://e.lanbook.com/reader/journalArticle/595721> (In Russ)
10. Yarmolenko A.S., Skobenko O.V.: Application of the theory of wavelets in the compression and filtering of geoinformation. Notes of the Mining Institute 234, 612-623 (2018). <https://e.lanbook.com/reader/journalArticle/71014> (In Russ)
11. Suchkova L.I., Hussein H.M., Yakunin M.A., Yakunin A.G.: Scalable software and hardware solutions for environmental and technical monitoring. Polzunovskiy Almanac 1, 75-82 (2013). <https://e.lanbook.com/reader/journalArticle/379151> (In Russ)
12. Hussein H.M., Kuntz R.V., Suchkova L.I., Yakunin A.G.: Design and implementation of weather and technology process monitoring systems. Herald of the Altai State University 1-1(77), 210-214 (2013).
13. Yakunin A.G., Hussein H.M.: Hardware-software and algorithmic provision of multipoint systems for long-term monitoring of dynamic processes. IOP Conf. Series: Journal of Physics 881 012028 (2017). DOI: 90.1088/1742-6596/881/1/012028
14. Hussein H.M., Suchkova L.I., Yakunin A.G.: Ways for improving methods of data storing in monitoring systems. Polzunovskiy Almanac 2, 2012; 48-50
15. Hussein, H. M., Yakunin A.G.: Data Differencing Method To Optimize Data Storing In Weather Monitoring System. Polzunovskiy Herald 2013; 2:65-68 https://journal.altst.ru/media/f/old2/pv2013_02/pdf/065hussein.pdf
16. Melnikov E.M., Yakunin A.G.: Investigation of the effectiveness of lossless compression methods for temperature monitoring data. Proceedings of the XX International scientific and technical conference "Measurement, control, informatization" 32-38 (2019). <https://elibrary.ru/item.asp?id=41166261> (In Russ)
17. Hussein H. M., Yakunin A.G.: Simple Curve Smoothing Methods For Weather Monitoring System. Materiali XI Mezinarodni vedecko-practicka conference «Veda a Technologie: krok do budoucnosti-2015», Dil 17. Technické vedy. Praha, 2015; 73-76.
18. Hussein H., Yakunin A., Suchkova L. A Comparison of Data Compression Methods for Solving Problems of Temperature Monitoring. MATEC Web of Conferences, 79 (2016). <http://dx.doi.org/10.1051/mateconf/20167901076>.
19. Zhen Ch., Ren : Design and Realization of Data Compression in Real-Time Data-base." *IEEE International Conference on Computational Intelligence and Software Engineering*, China, 11 – 13 Dec 2009: 340-243
20. Ray G.: Data Compression in Databases. Master's Thesis, Dept. of Computer Science and Automation, Indian Institute of Science, June (1995).
21. Yang H.: On the performance of data compression algorithms based upon string matching. *IEEE Transactions on Information Theory*. 1998; 44(1):47- 65.
22. Lin M., Chang Y. Yi: A New Architecture of a Two-Stage Lossless Data Compression and Decompression Algorithm. *IEEE transactions on VLSI* 2009; 17(9): 1297- 1303.
23. Tamrakar A., Nanda V. A Compression Algorithm for Optimization of Storage Consumption of Non Oracle Database 2012; 1(5):39-43.
24. Yakunin A. and Suchkova L. Application of the E – layer model for solving the problems of parametric estimation in measuring devices. MATEC Web of Conferences 155, 01020 (2018) DOI: <https://doi.org/10.1051/mateconf/201815501020>
25. Suchkova L.I., Yakunin A.G.: Application of E-Regions of the Parameters of the Model Function of the Signal for the Extraordinary Situations Detection. 2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon, 3-4 Oct. 2018). *IEEE Xplore* 1-6 (2019), URL: <https://ieeexplore.ieee.org/document/8602766>, DOI: 10.1109/FarEastCon.2018.8602766
26. Suslova S.V., Yakunin A.G. A method for extracting a trend of a non-stationary process with adaptation of approximation intervals. Patent of Russia № 2645273. 2018; 5 (In Russ)

Сведения об авторе:

Якунин Алексей Григорьевич, доктор технических наук, профессор, заведующий кафедрой информатики, вычислительной техники и информационной безопасности; almpas@list.ru,

Information about the author:

Aleksey G. Yakunin. Dr. Sci. (Eng), Prof., Head of the Department of Informatics, Computer engineering and Information security; almpas@list.ru,

Конфликт интересов/Conflict of interest.

Автор заявляет об отсутствии конфликта интересов/The author declare no conflict of interest.

Поступила в редакцию/Received 18.07.2023.

Одобрена после рецензирования/ Revised 20.08.2023.

Принята в печать/Accepted for publication 20.08.2023.