

На правах рукописи

Соколов Игорь Сергеевич

**Методы и модели восстановления структуры
группового телеметрического сигнала**

05.13.18 – Математическое моделирование, численные методы и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Работа выполнена в *Санкт-Петербургском государственном электротехническом университете «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ) на кафедре «Математическое обеспечение и применение ЭВМ».*

Научный руководитель: *кандидат технических наук,
доцент,
Балтрашевич Владимир Эдуардович*

Официальные оппоненты: *доктор технических наук,
профессор,
Куприянов Михаил Степанович,
декан факультета компьютерных технологий и информатики, СПбГЭТУ*

*кандидат технических наук,
Игнатов Дмитрий Игоревич,
старший преподаватель кафедры анализа данных и искусственного интеллекта,
НИУ ВШЭ*

Ведущая организация: *Открытое акционерное общество «Государственный ракетный центр имени академика В.П. Макеева»*

Защита состоится «16» мая 2012 г. в 15 часов на заседании совета по защите докторских и кандидатских диссертаций Д212.238.01 при СПбГЭТУ, расположенном по адресу: 197376, Россия, Санкт-Петербург, улица Профессора Попова, дом 5

С диссертацией можно ознакомиться в библиотеке СПбГЭТУ «ЛЭТИ».

Автореферат разослан «13» апреля 2012 г.

Ученый секретарь
совета по защите докторских и
кандидатских диссертаций Д212.238.01,
к.т.н.

Щеголева Н.Л.

Общая характеристика работы

Актуальность работы. Современный этап развития телеметрических систем характеризуется быстрым ростом объема передаваемой информации. В настоящее время для ракетно-космической техники (РКТ) количество передаваемых параметров исчисляется тысячами, а длины временных рядов, определяющих поведение параметров, превышают миллион отсчетов. Наряду с объемом передаваемой информации изменилась и сложность передаваемого потока телеметрии: с целью оптимального использования ограниченных по пропускной способности каналов связи стали применяться разнообразные схемы коммутации измерительной информации. В связи с этим возникает задача контроля и восстановления структуры передаваемых информационных потоков. В рамках этой задачи, как правило, предполагается, что структура анализируемого потока неизвестна, и ее необходимо восстановить.

В работе рассматривается ограниченный класс телеметрических потоков, которые обладают сложной циклической структурой и поступают с объектов РКТ. В ракетно-космической отрасли такой поток принято называть *групповым телеметрическим сигналом* (ГТС). Частичные знания о структуре каждого ГТС известны и регламентируются стандартом, который определяет общие правила формирования сигнала. Полная информация о структуре является для каждого объекта РКТ частной и выбирается в соответствии с целями и задачами испытаний и штатной эксплуатации.

Существует определенный задел в литературе, представленный работами Балтрашевича В. Э., Васильева А. В., Кукушкина С. С. и опытно-конструкторскими работами, проводимыми в научно-исследовательском центре представления и контроля информации по космодрому Плесецк, однако, большинство из существующих решений находятся на этапе развития, когда большая часть действий выполняется непосредственно экспертом. При этом практически не используется информация о ранее накопленных ГТС с известными структурами. Как следствие, восстановление структуры одного ГТС с неизвестной структурой от объекта РКТ в среднем требует не менее нескольких часов. Необходимо подчеркнуть, что это труд высококвалифицированных специалистов, нехватка которых является острой проблемой. Таким образом, является актуальной задача разработки методов и моделей восстановления структуры ГТС, которые позволят сократить время восстановления.

Цель и задачи исследования — разработка методов и моделей, обеспечивающих сокращение времени восстановления структуры ГТС с неизвестной структурой за счет повышения степени автоматизации восстановления и использования информации о ранее накопленных ГТС с известными структурами.

Для достижения поставленной цели были решены следующие задачи:

- 1) анализ существующих решений задачи восстановления структуры ГТС

с неизвестной структурой и выработка путей преодоления их основных недостатков;

2) разработка трехэтапной процедуры восстановления структуры ГТС с неизвестной структурой;

3) разработка метода экспресс-поиска структуры ГТС с использованием информации о ранее накопленных ГТС с известными структурами;

4) разработка метода последовательного восстановления структуры ГТС с неизвестной структурой, базирующегося на анализе бинарного представления ГТС и позволяющего определить типы медленноменяющихся телеметрических параметров;

5) разработка графовой модели ГТС, учитывающей структуру ГТС и поведение входящих в его состав параметров, и метода комплексного поиска структуры ГТС, использующего данную модель;

6) программная реализации предлагаемой трехэтапной процедуры восстановления структуры ГТС.

Объектом исследования диссертационной работы является процесс восстановления структуры ГТС.

Предметом изучения являются методы и алгоритмы, позволяющие осуществить восстановление структуры ГТС, а также модели описания ГТС и его структуры.

Методы исследования. Теоретическая часть работы выполнена на основе численных методов математической статистики, машинного обучения, теории графов и алгоритмов на строках. В экспериментальной части работы применяются численные методы и методы имитационного моделирования. Для выполнения экспериментальной части создан и зарегистрирован в реестре программ для ЭВМ программный комплекс.

Научную новизну работы составляют следующие положения:

1) Трехэтапная процедура восстановления структуры ГТС с неизвестной структурой, включающая этап экспресс-анализа (применяется метод экспресс-поиска), этап полного анализа (метод последовательного восстановления) и этап уточнения (метод комплексного поиска).

2) Метод экспресс-поиска структуры ГТС с неизвестной структурой на базе аппроксимации распределения информационных слов ГТС. Метод позволяет для анализируемого ГТС при минимальном количестве априорной информации об его структуре определить набор потенциальных структур, представленных среди ранее накопленных ГТС с известными структурами.

3) Метод последовательного восстановления структуры ГТС с неизвестной структурой, позволяющий на основе анализа бинарного представления ГТС в автоматическом режиме восстановить его структуру и метод включает группу алгоритмов на основе корреляционного анализа и анализа характеристик изменения слов в кадре для построения модели формата кадра и алгоритм определения типов медленноменяющихся телеметрических параметров с использо-

ванием метрики неточного сравнения символьного представления временных рядов.

4) Графовая модель ГТС на базе атрибутированного графа отношений, учитывающая как структуру ГТС, так и поведение его составляющих параметров. Метод комплексного поиска структуры ГТС с неизвестной структурой на базе сравнения графовых моделей ГТС с использованием редакционного расстояния. Метод позволяет с целью уточнения структуры анализируемого ГТС определить наиболее схожие ГТС среди ранее накопленных с известными структурами.

Практическая значимость работы заключается в разработке алгоритмов и программ восстановления структуры ГТС, позволяющих сократить время, необходимое на восстановление структуры, в среднем в 4 раза.

Материалы диссертационной работы были использованы при выполнении опытно-конструкторской работы (ОКР), посвященной обработке измерительной информации РКТ и их оснащения (шифр «Модернизация»), и при выполнении ОКР, посвященной системе комплексного анализа результатов пусков РКТ (шифр «Математика-ПИК»), проводимых в ОАО «Научно-инженерный центр Санкт-Петербургского электротехнического университета» («НИЦ СПб ЭТУ»). Результаты работы внедрены в в/ч 85487 (ОКР «Математика-ПИК») и в в/ч 13991-П (ОКР «Модернизация»).

Работа поддержана программой в рамках конкурсов научных достижений аспирантов СПбГЭТУ «ЛЭТИ» в 2011 году.

Научные положения, выносимые на защиту:

- 1) трехэтапная процедура восстановления структуры ГТС;
- 2) метод экспресс-поиска структуры ГТС;
- 3) метод последовательного восстановления структуры ГТС, включающий набор алгоритмов построения модели формата кадра ГТС и алгоритм определения типов медленноменяющихся телеметрических параметров;
- 4) графовая модель ГТС и метод комплексного поиска структуры ГТС;

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались со специалистами ОАО «НИЦ СПб ЭТУ», на кафедре математического обеспечения и применения ЭВМ СПбГЭТУ, а также на следующих конференциях:

- 14-я всероссийская конференция «Математические методы распознавания образов» (ММРО-14), Владимирская обл., г. Суздаль, 2009 г.
- двенадцатая национальная конференция по искусственному интеллекту с международным участием (КИИ-2010), г. Тверь, 2010 г.
- XIV международная конференция по мягким вычислениям и измерениям (SCM'2011), Санкт-Петербург, 2011 г.
- научные сессии Национального исследовательского ядерного университета (НИЯУ) «МИФИ», Москва, 2010, 2011 и 2012 гг.
- научно-технические конференции профессорско-преподавательского

состава СПбГЭТУ, Санкт-Петербург, 2010, 2011 и 2012 гг.

Публикации. Материалы диссертации опубликованы в 11 печатных работах, из них 2 статьи в рецензируемых журналах, 1 статья в нерцензируемых журналах и 7 докладов в сборниках трудов конференций.

Разработанный автором комплекс анализа структур данных был зарегистрирован в реестре программ для ЭВМ (свидетельство № 2011616384).

Структура и объем диссертации. Диссертация состоит из введения, 4 глав, заключения и библиографии. Общий объем диссертации 147 страниц, из них 134 страниц текста, включая 36 рисунков. Библиография включает 99 наименований на 13 страницах.

Содержание работы

Во введении обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана практическая значимость полученных результатов, представлены выносимые на защиту научные положения.

В первой главе введены основные понятия предметной области, приведена постановка задачи восстановления структуры ГТС и осуществлен обзор существующих решений с указанием основных недостатков.

Под *телеметрическим параметром* (ТМП) понимается измерительная информация о поведении физического процесса, поведение которого подлежит измерению или контролю телеметрической системой. В зависимости от скорости изменения во времени ТМП делятся на *медленноменяющиеся параметры* (ММП) и *быстроменяющиеся параметры* (БМП). Первые характеризуются шириной спектра от 0 до 50 Гц, а вторые — от 1 до 100 кГц. Большинство параметров, передаваемых от объекта РКТ, относятся к классу ММП, поэтому в настоящей работе будет рассматриваться только этот класс.

Вводится понятие *типа ММП* как элемента из конечного алфавита априори заданных типов \mathcal{T} . С каждым типом ММП ассоциировано эталонное поведение \mathcal{E} , которое задается с помощью одной или нескольких реализаций:

$$\mathcal{E} = \{E_1, \dots, E_m\}, \quad (1)$$

где E_i — реализация, задаваемая временным рядом (e_1^i, \dots, e_n^i) и определяющая один из возможных вариантов поведения, m — количество реализаций.

В работе рассматривается ГТС как ограниченный класс телеметрических информационных потоков со сложной циклической структурой, поступающих с объектов РКТ. Межведомственной комиссией по измерительным средствам был разработан стандарт, регламентирующий особенности формирования ГТС — *IRIG Standart 106*.

В соответствие со стандартом, ГТС — структурированная последовательность информационных слов, получаемая в результате объединения зарегистрированных измерений ТМП на основе временного разделения каналов. Измерения ТМП передаются в виде *информационных слов*, длина в битах которых одинакова и определяется разрядностью установленных двоичных аналого-цифровых преобразователей. Один цикл опроса главным коммутатором всех непосредственно подключенных к нему датчиков порождает последовательность информационных слов, называемую *кадром*. *Суперкоммутация* и *субкоммутация* определяются как передачи данных с частотой, которая кратна (суперкоммутация) и субкратна (субкоммутация) частоте кадра. *Кратность суперкоммутации* показывает, сколько раз измерения данного параметра встречаются в одном кадре, а *глубина субкоммутации* — сколько параметров передается в одном слове кадра.

Работа объекта РКТ сопровождается переключениями режимов работы, что отражается в изменении схемы коммутации датчиков. Данный процесс изменения называется *сменой формата кадра*. В качестве примера переключения можно привести отделение ступеней ракеты.

Процесс распаковки ГТС на отдельные параметры называется *декоммутацией* и определяется как оператор:

$$D : (\mathcal{B}, \mathcal{S}) \rightarrow \mathcal{P}, \quad (2)$$

где $\mathcal{B} = \langle B_1, \dots, B_N \rangle$ — последовательность бит (бинарное представление ГТС), \mathcal{S} — структура ГТС, $\mathcal{P} = \{P_1, \dots, P_K\}$ — набор типизированных ТМП. Последний является кортежем:

$$P_i = \langle t_{P_i}, \mathbf{P}_i \rangle, \quad (3)$$

где $t_{P_i} \in \mathcal{T}$ — тип i -го ТМП, а $\mathbf{P}_i = (p_1^i, \dots, p_n^i)$ — временной ряд измерений ТМП.

Описание структуры \mathcal{S} состоит из информации о структуре коммутации датчиков в объекте (*формат кадра*) и информации о функциональном назначении каждого из передаваемых параметров (*информация о типах параметров*). Под *архивом ГТС* будем понимать набор исходных ГТС и соответствующих им описаний структур.

Задача восстановления структуры ГТС с неизвестной структурой: необходимо проанализировать бинарное представление ГТС \mathcal{B} и на основании знаний об общих принципах формирования сигнала, которые регламентируются стандартом IRIG Standart 106, а также информации о ранее накопленных ГТС с известными структурами, выявить такую структуру анализируемого ГТС \mathcal{S} , чтобы в результате его декоммутации, определенной в формуле (2), был получен истинный набор ТМП \mathcal{P} .

Основные существующие подходы к восстановлению структуры ГТС: графический подход на базе когнитивной графики и подход на основе анализа кор-

реляционных функций и характеристик изменения значений информационных слов в кадре.

В рамках графического подхода последовательность кадров визуализируется как строки матрицы, элементы которой в зависимости от значения битов являются черными или белыми штрихами. Особенности оцифровки аналоговой информации обуславливают большую изменчивость младших разрядов информационного слова по сравнению со старшими, как следствие, при корректных длинах кадра и информационного слова большинство столбцов либо не изменяются, либо можно визуально обнаружить закономерности изменений. Такой визуальный подход не приводит к гарантированно корректному результату и всецело задействует опыт анализатора.

В рамках второго подхода осуществляется обработка \mathcal{B} , сведенного в последовательность информационных слов $\langle X_i \rangle$ определенной разрядности L_c . Задача определения длины информационного слова и длины кадра решается путем оценки экспертом периодов между пиками *автокорреляционной функции* (АКФ). К числу недостатков относятся: большая часть операций выполняется экспертом, в соответствии с проведенными автором исследованиями только на основании анализа максимальных значений АКФ нельзя выполнить автоматическое определение длины кадра и длины слова. Для поиска суб- и суперкоммутаций производится анализ изменения слов в последовательности кадров, представленных в виде матрицы MT из L_{gts} строк (длина ГТС в кадрах) с одинаковым количеством столбцов в размере L_f информационных слов (длина кадра в словах). Предполагается оценивание экспертом математического ожидания изменения предшествующих и последующих по времени элементов матрицы MT : большие величины указывают на наличие субкоммутации. К числу недостатков относятся: большой объем операций, выполняемых непосредственно экспертом, в соответствии с проведенными автором исследованиями крайне высокая чувствительность в случае длин слов большой разрядности, наличие сбойных участков может сильно исказить вычисляемую статистику, что приводит к ложным выявлением коммутаций.

Указанные подходы не осуществляют определение типов ТМП, а следовательно, они не решают задачу восстановления структуры ГТС в полном объеме.

Во второй главе описывается предлагаемая трехэтапная процедура восстановления структуры ГТС с неизвестной структурой и вводятся предлагаемые модели. На рисунке 1 представлена схема данной процедуры, которая состоит из трех этапов:

- 1) Экспресс-анализ. При минимальном количестве априорной информации о структуре анализируемого ГТС определяются ближайшие ГТС в архиве ГТС (*метод экспресс-поиска*), структуры этих ГТС экспертом проверяются на применимость к анализируемому ГТС. Если эксперт решает, что ни одна из найденных структур не подходит, осуществляется следующий этап, иначе структура ГТС считается найденной.

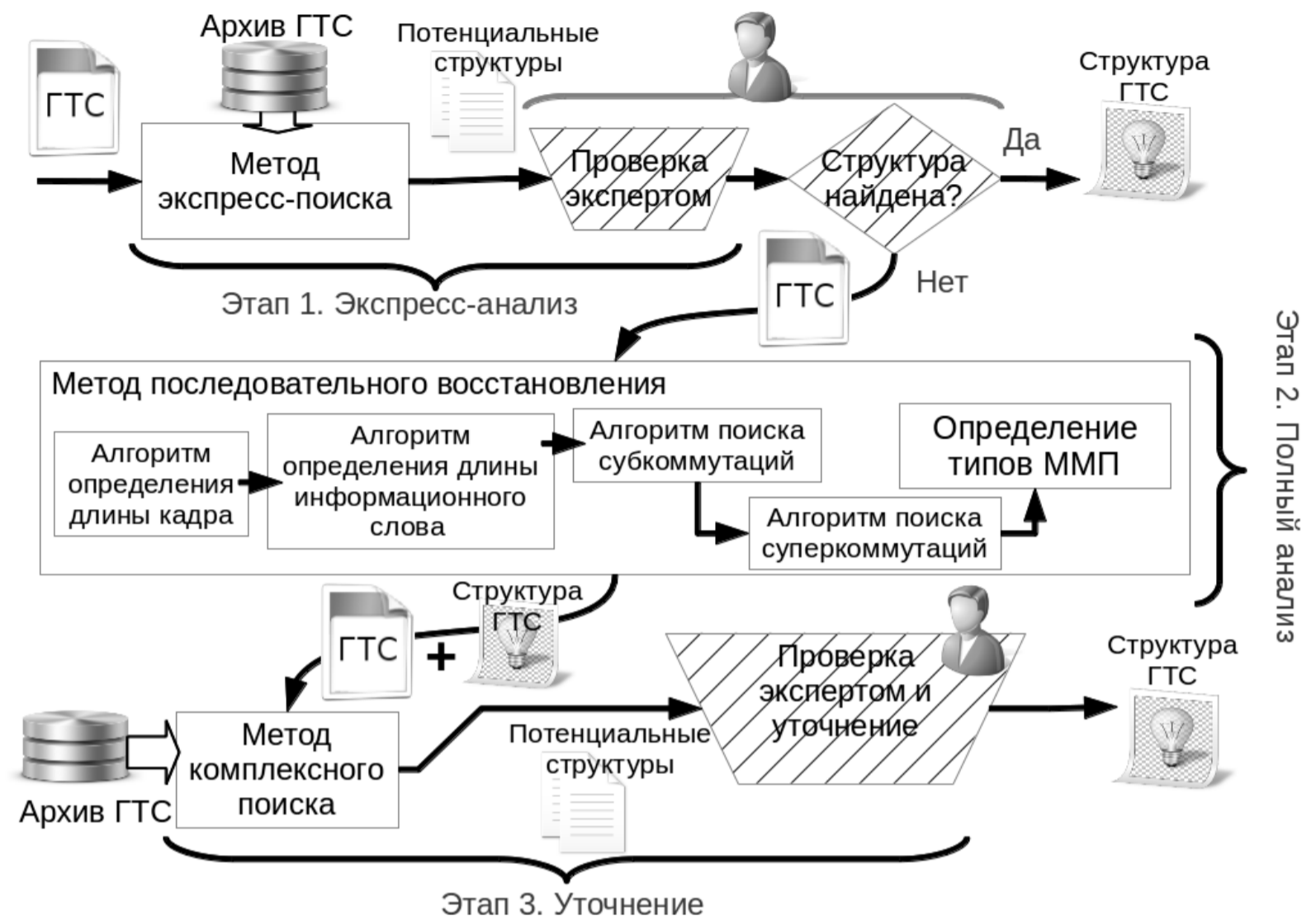


Рисунок 1. Схема трехэтапной процедуры восстановления структуры ГТС.

2) Полный анализ. Последовательно применяются алгоритмы для определения всех элементов модели формата кадра и определяются типы ММП (*метод последовательного восстановления*).

3) Уточнение. Осуществляется поиск в архиве ГТС схожих сигналов с учетом полученной структуры (*метод комплексного поиска*). Эксперт проверяет структуры найденных ГТС и уточняет изначально полученную структуру.

В основе методов поиска в архиве ГТС лежит тот факт, что в ракетно-космической отрасли структуры ГТС, полученные от одного и того же объекта РКТ, одинаковы или имеют минимальные различия. Как следствие, сходство ГТС свидетельствует о сходстве их внутренних структур.

С целью формализации описания формата кадра как схемы коммутации датчиков в ГТС вводится *модель формата кадра*:

$$F = \langle L_w, L_f, Sub, Sup \rangle, \quad (4)$$

где L_w — длина информационного слова в битах, L_f — длина кадра в словах, Sub — множество субкоммутиаций, Sup — множество суперкоммутиаций.

Множество суперкоммутиаций есть множество пар $Sup = \{(i, f)\}_{0...k}$, где i — индекс слова в кадре, f — кратность суперкоммутиации, k — количество суперкоммутиаций в кадре. *Множество субкоммутиаций* есть множество пар —

индекс слова в кадре и глубина субкоммутации $Sub = \{(j, d)\}_{0\dots m}$, где j — индекс слова в кадре, d — глубина субкоммутации, m — количество субкоммутаций в кадре.

Для того, чтобы учитывать при сравнении ГТС как информацию о поведении ТМП, так и структуру ГТС, была разработана графовая модель ГТС. Модель основана на атрибутированном графе отношений (attributed relational graph). Пусть M_V, M_E — это атрибуты вершин и ребер, соответственно, каждый атрибут имеет вид (σ, φ) , где σ — тип атрибута, $\varphi = (\varphi_1, \dots, \varphi_n)$ — вектор характеристик, ассоциированный с σ . Тогда *атрибутированный граф отношений* над множеством атрибутов $M = M_V \cup M_E$ есть:

$$G = \langle V, E, \mu, e \rangle, \quad (5)$$

где V — множество вершин, $E \subseteq V \times V$ — множество ребер, $\mu : V \rightarrow M_V$ — функция назначения атрибутов вершинам, $e : E \rightarrow M_E$ — функция назначения атрибутов ребрам.

В случае описания ГТС атрибутированный граф — это схема соединений датчиков и коммутаторов в объекте РКТ, где узлами являются вершины, описывающие субкоммутаторы, а конечные узлы — это параметры, измеряемые датчиками. Выделяется три типа атрибутов вершин: $T_V = \{fp, sp, com\}$, где fp — тип атрибута для БМП, sp — тип атрибута для ММП, com — тип атрибута коммутатора.

Атрибутом для коммутаторов является кортеж имен вершин–потомков: $P_{com} = \langle n_{v_1}, \dots, n_{v_m} \rangle$, где n_{v_1}, \dots, n_{v_m} — имена вершин v_1, \dots, v_m , связанных с коммутатором c : $(c, v_m) \in E, i = 1, \dots, m$. В зависимости от того, относится ли ТМП к классу БМП или ММП, атрибуты вершин задаются в спектральном или символьном представлении как $P_{fp} = \langle f_1, \dots, f_l \rangle$ и $P_{sp} = \langle s_1, \dots, s_w \rangle$ соответственно, где f_i — величина i -го коэффициента преобразования Фурье, l — количество рассматриваемых коэффициентов, s_i — i -й элемент символьного представления ММП, w — длина символьного представления.

Множество атрибутов ребер определяется как $T_E = \{l, r\}$, где l — атрибут ребер «непосредственно связан», r — атрибут ребер «элемент заменяется на», который используется для указания смены формата кадра. Указанные атрибуты ребер не имеют векторов характеристик.

Графовая модель ГТС есть кортеж:

$$\mathcal{G} = \langle V, E, \alpha_V, \alpha_E, L_w \rangle, \quad (6)$$

где V — датчики и коммутаторы, E — линии связей датчиков и коммутаторов, α_V — функция назначения вершинам соответствующих атрибутов $\{(com, P_{com}), (fp, P_{fp}), (sp, P_{sp})\}$, α_E — функция назначения ребрам атрибутов $\{l, r\}$, L_w — длина информационного слова.

В третьей главе приводится описание предлагаемых методов восстановления структуры ГТС.

Метод экспресс-поиска направлен на максимально быстрое получение потенциальных структур ГТС за счет сравнения характеристик анализируемого ГТС с неизвестной структурой с характеристиками ранее накопленных ГТС, для которых известны структуры. При этом набор характеристик ГТС строится только на основании бинарного представления ГТС и длины информационного слова.

Для текстов на естественном языке известен закон Ципфа – Мандельброта: «произведение ранга некоторого слова и частоты его встречаемости в тексте является величиной постоянной, имеющей примерно одинаковое значение для любого слова из словаря рассматриваемого текста и характерной для данного языка». Дальнейшие исследования показали, что закону Ципфа – Мандельброта удовлетворяют не только слова из текстов на естественном языке, но и практически все объекты современного информационного пространства. Исходя из этого, в качестве вектора характеристик предлагается использовать коэффициенты C , γ , d аппроксимирующей функции частотно-рангового распределения значений слов:

$$\Phi(r) = C \cdot r^{-\gamma \cdot \exp(d \cdot r)}, \quad (7)$$

где $\Phi(r)$ — относительная частота информационного значения слова с рангом r , γ — константа, определяемая свойствами анализируемого ГТС, C — константа, определяемая объемом выборки, d — параметр увеличения коэффициента частотно-рангового соотношения. Процесс подбора параметров C , γ , d является задачей оптимизации с целевой функцией:

$$\sum_{r=1}^N (f_r - \Phi(r))^2 \rightarrow \min_{C, \gamma, d}, \quad (8)$$

где f_r — частота встречаемости значения с рангом r .

Применение предложенного метода позволяет в сокращенные сроки получить приближенный набор потенциальных структур ГТС. Потенциальный недостаток метода связан с большим размером словаря, который в худшем случае составляет 2^{L_w} слов.

Метод последовательного восстановления направлен на автоматическое получение формата кадра за счет анализа бинарного представления ГТС \mathcal{B} с использованием знаний о принципах формирования ГТС и осуществляет последующее определение типов ТМП. В рамках метода выделяются два подэтапа: восстановление формата кадра и определение типов ММП. Процесс восстановления формата кадра состоит из последовательного применения разработанных алгоритмов, осуществляющих определение одного из элементов модели формата кадра F в формуле (4).

Алгоритм определения длины кадра является модификацией существующего алгоритма и базируется на анализе АКФ при различных длинах информационного слова. Благодаря предложенному автором критерию наблюдаемости

маркера алгоритм автоматизирован. Критерий основан на том, что в начале кадра всегда находится маркер — специальный ТМП, значение которого не изменяется на протяжении всего сеанса телеметрирования.

Алгоритм определения длины информационного слова осуществляет корреляционный анализ частоты изменения бит в словах и основан на том, что для функциональных параметров наиболее часто изменяемыми являются младшие биты двоичного представления.

Алгоритмы поиска суб- и суперкоммутаций являются модификацией существующих алгоритмов. *Алгоритм поиска субкоммутаций* основан на том, что в результате объединения в субкоммутаторе измерений различных параметров среднее изменение значения слова с субкоммутацией будет иметь достаточно большое абсолютное значение. Дополнительно, глубина субкоммутации равна периоду АКФ, построенной для измерений анализируемого слова. Для определения периода автором был разработан алгоритм, в рамках которого осуществляется итеративный анализ упорядоченного множества пиков АКФ на наличие периода. *Алгоритм поиска суперкоммутации* строит иерархию частей кадров, соответствующих возможным кратностям суперкоммутации, и последовательно анализирует их характеристики изменения значений информационных слов. За кратность суперкоммутации принимается значение, при котором следующее значение кратности приводит статистику изменения анализируемого слова к резкому скачку по значению. Такой подход позволяет снизить чувствительность оригинального алгоритма в случае длин слов большой разрядности. Дополнительно применяемое отсечение множества потенциальных кратностей для анализируемого слова сокращает объем вычислений.

Определение типа ММП для анализируемого временного ряда измерений строится на базе поиска ближайших реализаций среди заданных эталонов типов $\mathcal{E}_\alpha, \dots, \mathcal{E}_\omega$. Метрика различия временных рядов строится на преобразовании исходного ряда в символьное представление, к которому применяется алгоритм вычисления редакционного расстояния. Данное решение учитывает кусочно-постоянную модель поведения параметров. В рамках нее поведение ММП является последовательностью участков, на каждом из которых сохраняется своя модель поведения, и продолжительность этих этапов может варьироваться. Применяемое редакционное расстояние позволяет снизить «стоимость» несущественных отклонений в поведении параметров и учесть различную продолжительность этапов.

Алгоритм построения символьного представления временного ряда является модификацией существующего алгоритма SAX (Symbolic Aggregate approximation). В алгоритме для исходного временного ряда $C = (c_1, \dots, c_N)$ с целью сокращения размерности строится кусочно-константная аппроксимация (ККА) $\bar{C} = (\bar{c}_1, \dots, \bar{c}_W)$, где N — длина исходного временного ряда, а W — количество сегментов аппроксимации. Далее, осуществляется нормализация к виду: $\bar{c}_i \in [-1, 1], \forall i = 1, \dots, W$. Для уменьшения влияния выбросов осуществляется

предварительная очистка и вычисление максимального (минимального) значения как медианы среди максимальных (минимальных) значений. Перевод из ККА в символьное представление производится на основании уровней, которые задаются как отсортированный список чисел $B = (\beta_1, \dots, \beta_{K-1})$, распределенных в отличие от оригинального алгоритма SAX равномерно на множестве $[-1, 1]$, где K — размер алфавита. Каждому отсчету \bar{c}_i назначается буква алфавита: $a_j \leftrightarrow \beta_{j-1} < \bar{c}_i \leq \beta_j$. В результате получается строка символов $S = (s_1, \dots, s_W)$, где $s_i \in \{a_0, \dots, a_K\}$.

Для сравнения символьных представлений параметров предлагается использовать *редакционное расстояние* (или *расстояние редактирования*) — взвешенная сумма минимального количества операций редактирования (вставки, удаления, или замены одного символа на другой), необходимых для превращения одной строки в другую. Формально расстояние редактирования:

$$d(S_1, S_2) = \min_{(e_1, \dots, e_k) \in Y(S_1, S_2)} \sum_{i=1}^k C(m_i), \quad (9)$$

где $Y(S_1, S_2)$ — множество путей редактирования для трансформации строки S_1 в S_2 , и $C(m_i)$ — функция стоимости операции m_i . Цены операций зависят как от вида операции, так и от участвующих в ней символов. В работе задачу для произвольных весов решает алгоритм Вагнера – Фишера.

Метод комплексного поиска осуществляет поиск ближайших ГТС в архиве с учетом предварительно определенной структуры и позволяет эксперту уточнить изначально полученную структуру. Сравнение ГТС осуществляется в рамках введенной ранее графовой модели ГТС. Для нахождения ближайших ГТС используется расстояние редактирования на графах (*graph edit distance*). Такое расстояние позволяет уменьшить степень влияния возможных ошибок в первоначальной структуре при сравнении моделей.

Ключевая идея расстояния редактирования на графах заключается в определении набора операций редактирования графов, которые могут быть применены, чтобы привести два графа к единому виду. Набор операций редактирования определен как: вставка, удаление и изменение, как узлов, так и ребер. Для каждой операции вводится стоимость, зависящая как от вида операции, так и от участвующих в ней элементов. Пусть \mathcal{G}_1 и \mathcal{G}_2 — это графовые модели ГТС, тогда расстояние редактирования между ними:

$$d(\mathcal{G}_1, \mathcal{G}_2) = \min_{(e_1, \dots, e_k) \in Y(\mathcal{G}_1, \mathcal{G}_2)} \sum_{i=1}^k \mathcal{H}(m_i), \quad (10)$$

где $Y(\mathcal{G}_1, \mathcal{G}_2)$ — множество путей редактирования для трансформации \mathcal{G}_1 в \mathcal{G}_2 , и $\mathcal{H}(m_i)$ — функция стоимости операции m_i .

Вычисление редакционного расстояния для графов осуществляется с помощью алгоритма деревьев поиска, который осуществляет поиск в простран-

стве всех возможных отображений узлов и ребер первого графа в узлы и ребра второго графа. Для произвольного графа время работы алгоритма экспоненциально зависит от количества узлов, поэтому такой алгоритм неприменим. Автором предлагается использовать особенность графовой модели ГТС, заключающуюся в «слоистой» структуре графа (один слой — это один формат кадра) для построения эффективного алгоритма сравнения. В предлагаемом алгоритме происходит попарное сравнение всех форматов кадров, являющихся деревьями, одной графовой модели с другой. После этого решается задача о назначении с целью минимизации суммы расстояний и идет подсчет итогового расстояния редактирования для оптимального назначения, включающего стоимости ребер для смен формата кадров. Сложность предлагаемого алгоритма $O(\max(N_1, N_2)^2 \cdot L_f)$, где N_1, N_2 — количество смен кадров в первой графовой модели и во второй, соответственно.

В четвертой главе приведены описания экспериментальных исследований, проведенных для методов и алгоритмов, разработанных в рамках настоящей диссертации.

Исследования проводились с использованием разработанного программного комплекса [3], реализованном на языке высокого уровня *Java*TM. Все замеры времени исполнения проводились на персональном компьютере с центральным процессором Intel®CoreTM i3-2310M с тактовой частотой 2.10 ГГц и объемом оперативной памяти 4096 Мбайт.

Для проведения экспериментальных исследований был выбран набор измерений реальных параметров (свыше 10000 временных рядов), количество сбойных участков не более 10%. Указанный набор содержал дискретные параметры (*DP*), медленноменяющиеся функциональные параметры, представленные рядом целых чисел (*SVP*), быстро меняющиеся функциональные параметров, представленные рядом целых чисел (*FSP*), параметры, представленные числами с плавающей точкой (*FPP*), и прочие (*OP*). Обозначим данный набор как $\mathcal{U} = \{U_1, \dots, U_N : U_i = (u_1^i, \dots, u_m^i)\}$, где u_j^i — это j -ое измерение i -го временного ряда, а N — это количество тестовых параметров. При этом для каждого временного ряда U определен T — тип передаваемого параметра: $f : U \rightarrow T$. На основании набора \mathcal{U} и генерации случайных структур, которые отвечают правилам формирования ГТС, был создан генератор модельных данных. В качестве стандартных настроек генератора использовались типовые для РКТ характеристики: $L_w \in \{8, 16\}$, $L_f \in [256, 384]$, $|Sub| = 20\%$, $|Sup| = 30\%$, $DP = 5\%$, $SVP = 55\%$, $FVP = 20\%$, $FPP = 10\%$, $OP = 10\%$.

Время работы алгоритмов представлено в таблицах 1 и 2. Средний процент правильно определенных элементов формата кадра представлен в таблице 3. Данные, полученные для предлагаемых алгоритмов, являются средними показателями по 50 экспериментов.

Сравнение предлагаемого алгоритма определения типов ММП проводилось с алгоритмом SAX и алгоритмом, построенным на базе быстрого преоб-

Таблица 1

Время работы в минутах алгоритмов определения длины кадра и слова.

Размер ГТС, Мбайт	50	100	200	400
Существующий алгоритм (длина слова и кадра)	34	35,1	37,6	41,3
Предлагаемый алгоритм определения длины кадра	0,2	0,4	0,8	1,6
Предлагаемый алгоритм определения длины слова	0,3	0,7	1,3	2,6
Суммарное время определения длины кадра и слова	0,5	1,1	2,1	4,2

Таблица 2

Время работы в минутах алгоритмов поиска суб- и суперкоммутаций.

Длина кадра, информационные слова	128	256	512	1024	2048
Существующий алгоритм поиска субкоммутаций	9,66	18,20	35,26	69,40	137,66
Существующий алгоритм поиска суперкоммутаций	26,67	53,33	106,67	213,33	426,67
Предлагаемый алгоритм определения субкоммутаций	0,75	1,67	4,21	8,56	17,24
Предлагаемый алгоритм определение суперкоммутаций	1,15	2,35	4,89	9,87	19,21

Таблица 3

Средний процент правильно определенных элементов модели формата кадра.

Элемент модели	L_f	L_w	Sub	Sup
Типовые	98	80	91,3	90,1
$ Sub > 60\%$	92	40	90,2	92,4
$ Sup > 60\%$	98	78	93,0	89,5
$DP = 5\%$, $SV\mathcal{P} = 35\%$, $FVP = 40\%$, $FPP = 10\%$, $OP = 10\%$	92	42	88,1	85,4
$DP = 5\%$, $SV\mathcal{P} = 35\%$, $FVP = 10\%$, $FPP = 40\%$, $OP = 10\%$	80	20	72,3	71,7

разования Фурье (БПФ) и меры Жаккара. Были выбраны 4 типа ММП, каждый представлен 20 реализациями: 16 использовалась для формирования эталона \mathcal{E} , 4 для тестирования. В общей сложности с различными комбинациями параметров было проведено: $6 \cdot \binom{16}{20}^2$ экспериментов и получен средний процент правильно определенных типов: БПФ = 85,5 %, SAX = 93,75 %, Предлагаемый = 96,4%.

В таблице 4 представлены результаты экспериментальных исследований метода экспресс-поиска (МЭП) и метода комплексного поиска (МКП) структуры ГТС. Строки соответствуют пяти экспериментам, проведенным с разными настройками генератора. Для каждого эксперимента был сгенерирован архив в 100 ГТС. Для экспресс-поиска приведен средний процент для 100 тестовых ГТС, для которых среди 10 найденных хотя бы один обладал правильной струк-

Процент корректных ближайших ГТС, определенных методами поиска.

Метод	МЭП	МКП
Стандартные	80	96
$ Sub > 60\%$	82	94
$ Sup > 60\%$	81	94
$DP = 5\%, SV\mathcal{P} = 35\%, FVP = 40\%, FPP = 10\%, OP = 10\%$	78	89
$DP = 5\%, SV\mathcal{P} = 35\%, FVP = 10\%, FPP = 40\%, OP = 10\%$	76	85

турой. Для комплексного поиска рассматривался только первый ближайший.

В рамках внедрения предлагаемые методы восстановления структуры ГТС проходили испытание на реальных данных. Общее число ГТС в архиве — 83, представлено 19 различных структур. Количество ГТС, для которых производилось восстановление структуры, составляло 60.

Для восстановления была применена предлагаемая процедура восстановления. После проведения этапа экспресс-анализа были подобраны структуры для 19 ГТС. Общее время этапа составило 1 час 12 минут. Применение метода последовательного восстановления для оставшихся ГТС потребовало около 13 часов. В результате применения комплексного поиска для 28 ГТС были найдены ГТС со схожими или точно совпадающими структурами, что позволило уточнить для 16 ГТС изначальную структуру. Общее время, которое потребовалось на поиск и уточнение, составило ≈ 6 часов. Общее время анализа для 60 ГТС составил ≈ 20 часов. По оценкам экспертов на анализ аналогичного числа ГТС без использования предлагаемых автором методов потребовалось бы свыше 90 часов работы в идеализированных условиях.

В заключении подводятся итоги работы, делаются выводы об эффективности и применимости полученных результатов, описываются дальнейшие пути развития работы.

Основные результаты работы

1) Разработана трехэтапная процедура восстановления структуры ГТС с неизвестной структурой, позволяющая для ускорения восстановления и уточнения структуры ГТС задействовать информацию о ранее накопленных групповых телеметрических сигналах с известной структурой. При экспериментальных исследованиях на реальных данных процедура позволила ускорить восстановление структуры ГТС в 4 раза.

2) Разработан метод экспресс-поиска структуры ГТС с неизвестной структурой на базе аппроксимации распределения информационных слов группового телеметрического сигнала. Метод позволяет для анализируемого ГТС при минимальном количестве априорной информации об его структуре определить набор потенциальных структур, представленных среди ранее накопленных ГТС

с известными структурами.

3) Разработан метод последовательного восстановления структуры ГТС с неизвестной структурой, позволяющий на основе анализа бинарного представления ГТС с неизвестной структурой в автоматическом режиме восстановить его структуру. Предложена модель формата кадра, позволяющая в автоматизированном режиме провести декоммутацию ГТС на набор отдельных параметров. В соответствии с экспериментальными исследованиями метод имеет средний процент правильно определяемых элементов модели формата кадра 79%. Предложен алгоритм определения типов ММП на базе символьного представления временных рядов, который за счет введения операций редактирования со стоимостями позволяет уменьшить влияние несущественных отклонений в поведении параметров. Для проведенных экспериментов предлагаемый алгоритм осуществляет определение типов с точностью примерно 95%.

4) На базе атрибутированного графа отношений разработана графовая модель ГТС, позволяющая учитывать как структуру, так и поведение составляющих его параметров. Разработан метод комплексного поиска структуры ГТС с неизвестной структурой на базе сравнения графовых моделей ГТС с использованием редакционного расстояния. В соответствии с экспериментальными исследованиями метод позволяет увеличить точность восстановления структуры в среднем до 90%.

5) Разработан программный комплекс, реализующий трехэтапную процедуру восстановления структуры ГТС. Данный комплекс лег в основу внедренных программных комплексов.

Список публикаций

В изданиях рекомендованных ВАК РФ:

1. Геппенер В.В., Горбачева И.В., Жукова Н.А., Соколов И.С. Идентификация телеметрических параметров с использованием нейронных сетей // Нейрокомпьютеры: разработка, применение. – 2009. – Т. 11. – С. 39–44.
2. Балтрашевич В.Э., Жукова Н.А., Соколов И.С. Графовая модель группового телеметрического сигнала со сменой кадра // Известия СПбГЭТУ «ЛЭТИ». – 2010. – № 10. – С. 12–17.

Свидетельства о государственной регистрации программ для ЭВМ:

3. Свидетельство о государственной регистрации программы для ЭВМ № 2011616384. Соколов И.С. «Комплекс анализа структур данных (КАСД)».

Статьи и материалы конференций:

4. Жукова Н.А., Соколов И.С. Метод восстановления структуры группового телеметрического сигнала на основе графовой модели // Труды СПИИРАН. – 2010. – № 13. – С. 45–66.
5. Балтрашевич В.Э., Васильев А.В., Жукова Н.А., Соколов И.С. Метод идентификации групповых телеметрических сигналов на основе частотно рангового распределения // 14-я всероссийская конф. мат. методы распознавания образов (ММРО-14): Сб. докл. – М.: МАКС Пресс, 2009. – С. 305–308.
6. Балтрашевич В.Э., Витол А.Д., Жукова Н.А., Соколов И.С. Интеллектуальный подход к анализу структуры и семантики ГТС // Труды 12-й нац. конф. по искусственному интеллекту с междунар. участием (КИИ-2010). – М.: Физматлит, 2010. – Т. 2. – С. 1-4.
7. Соколов И.С., Чех А.И. Метод формирования семантического описания структурированного потока измерений // Сб. докл. XIV междунар. конф. по мягким вычислениям и измерениям (SCM'2011). – СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2011. – Т. 2. – С. 25-27.
8. Балтрашевич В.Э., Соколов И.С. Применение графовой модели для представления структуры группового телеметрического сигнала // 62-й науч.-техн. конф. проф.-преп. состава университета: Сб. докл. студентов, аспирантов и молодых ученых. – СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2010. – С. 98-103.
9. Балтрашевич В.Э., Жукова Н.А., Соколов И.С. Определение структуры циклического информационного потока применительно к групповому телеметрическому сигналу // Науч. сессия НИЯУ МИФИ-2010: Аннот. докл. – М.: МИФИ, 2010. – Т. 3. – С. 75–76.
10. Соколов И.С. Метод построения графовой модели группового телеметрического сигнала // Науч. сессия НИЯУ МИФИ-2011: Аннот. докл. – М.: МИФИ, 2011. – Т. 3. – С. 77-78.
11. Соколов И.С. Метод неточного сравнения медленноменяющихся телеметрических параметров с использованием символьного представления // Науч. сессия НИЯУ МИФИ-2012: Аннот. докл. – М.: МИФИ, 2012. – Т. 3. – С. 89-90.