

ISSN 1818-7900 (Print)
ISSN 2410-0420 (Online)

N* Новосибирский
государственный
университет
***НАСТОЯЩАЯ НАУКА**

Вестник

ИТГУ

Научный журнал
Основан в ноябре 1999 года

2024
Том 22
№ 1

Серия: Информационные технологии

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ВЕСТНИК НГУ

Серия: Информационные технологии. 2024. Т. 22, № 1

Периодичность 4 раза в год
Выходит на русском языке
Индекс по каталогу Пресса России 11226

Журнал включен в перечень ведущих рецензируемых научных изданий
и журналов, рекомендуемых Высшей аттестационной комиссией
для публикации основных научных результатов диссертаций
на соискание ученой степени кандидата и доктора наук



Журнал включен
в Российский индекс научного цитирования

Электронная версия журнала размещена на платформе
Научной электронной библиотеки: www.elibrary.ru

Редакционная коллегия серии

ВрИО главного редактора д-р физ.-мат. наук, проф. *М. М. Лаврентьев*
Зам. главного редактора доцент *А. В. Авдеев*
Ответственный секретарь *Д. П. Иксанова*

акад. РАН И. В. Бычков (Иркутск), акад. РАН Н. А. Колчанов (Новосибирск)
акад. РАН Ю. И. Шокин (Новосибирск), чл.-корр. РАН С. И. Смагин (Хабаровск)
чл.-корр. РАН В. В. Шайдуров (Красноярск), проф. Б. М. Глинский (Новосибирск)
проф. Горбань А. Н. (Лестер, Великобритания), проф. Е. П. Гордов (Томск)
проф. Б. С. Добронев (Красноярск), проф. А. М. Елизаров (Казань)
проф. Г. Н. Ерохин (Калининград), проф. А. И. Камышников (Ханты-Мансийск)
проф. Г. П. Карев (Мариленд, США), проф. М. М. Лаврентьев (Новосибирск)
проф. В. Э. Малышкин (Новосибирск), проф. Н. Н. Миренков (Айзу, Япония)
проф. Н. М. Оскорбин (Барнаул), проф. Д. Е. Пальчунов (Новосибирск)
проф. Т. Пизанский (Любляна, Словения), проф. В. П. Потапов (Кемерово)
проф. О. И. Потатуркин (Новосибирск), проф. В. А. Серебряков (Москва)
проф. А. В. Старченко (Томск), проф. Д. А. Тусупов (Астана, Казахстан)

Адрес редколлегии
ул. Пирогова, 1, Новосибирск, 630090, Россия
Телефон: (383) 363 42 46
E-mail: inftech@vestnik.nsu.ru

Журнал зарегистрирован в ФС по надзору в сфере связи, информационных технологий
и массовых коммуникаций (свидетельство ПИ № ФС77-40143 от 04.06.2010)

Сдано в набор 19.04.2024. Подписано в печать 20.06.2024
Бумага офсетная № 1. Формат 60 × 84/8. Гарнитура Times New Roman
Печать цифровая. Усл. п. л. 8,83. Тираж 27 экз.
Заказ № 144

Отпечатано в Издательско-полиграфическом центре НГУ
ул. Пирогова, 2, Новосибирск, 630090, Россия

© Новосибирский государственный университет, 2024

ВЕСТНИК НОВОСИБИРСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

Научный журнал
Основан в ноябре 1999 года

Серия: Информационные технологии

2024. Том 22, № 1

СОДЕРЖАНИЕ

<i>Власенко А. Ю., Мичуров М. А., Царёв В. Д., Курбатов М. А.</i> Построение комплекса автоматизированной отладки фрагментированных программ.....	5
<i>Котлер В. Д., Платонова М. В., Климова Е. Г.</i> Структура информационно-вычислительной системы для решения задачи усвоения данных при моделировании окружающей среды.....	21
<i>Лобач Д. И.</i> О применении кластерного анализа для развития методов оценки безопасности технических систем	31
<i>Малахов С. В., Якупов Д. О., Осипова А. А., Копылова Д. А., Зеленина Е. А.</i> Применение системы массового обслуживания для исследования характеристик канала связи IoT-сетях	49
<i>Платонова М. В., Котлер В. Д., Климова Е. Г.</i> Оценка приземной концентрации метана на основе алгоритма ансамблевого фильтра Калмана с привлечением транспортной химической модели	62
Информация для авторов	73

V E S T N I K

NOVOSIBIRSK STATE UNIVERSITY

Scientific Journal
Since 1999, November
In Russian

Series: Information Technologies

2024. Volume 22, № 1

CONTENTS

<i>Vlasenko A. Y., Michurov M. A., Tsarev V. D., Kurbatov M. A.</i> Constructing a complex for automated debugging of fragmented programs.....	5
<i>Kotler V. D., Platonova M. V., Klimova E. G.</i> Structure of an information and computing system for solving the problem of data assimilation in environmental modeling.....	21
<i>Lobach D. J.</i> On the application of cluster analysis for the development of safety assessment methods for technical systems	31
<i>Malakhov S. V., Yakupov D. O., Osipova A. A., Kopylova D. A., Zelenina E. A.</i> The use of a queuing system to study the characteristics of the communication channel in IoT networks.....	49
<i>Platonova M. V., Kotler V. D., Klimova E. G.</i> Estimation of surface methane concentration based on the ensemble Kalman filter algorithm using a transport chemical model	62
Instructions for Contributors.....	73

Editor in Chief M. M. Lavrentiev

Vice-Editor A. V. Avdeev

Executive Secretary D. P. Iksanova

Editorial Board of the Series

- I. V. Bychkov*, professor, academician (Irkutsk), *B. M. Glinsky*, professor (Novosibirsk)
A. N. Gorban, professor (Lester, GB), *E. P. Gordov*, professor (Tomsk)
B. S. Dobronets, professor (Krasnoyarsk), *A. M. Elizarov*, professor (Kazan)
G. N. Erokhin, professor (Kaliningrad), *A. I. Kamyshnikov*, professor (Khanty-Mansijsk)
G. P. Karev, professor (Maryland, USA), *N. A. Kolchanov*, professor, academician (Novosibirsk)
M. M. Lavrentjev, professor (Novosibirsk), *V. E. Malyshkin*, professor (Novosibirsk)
N. N. Mirenikov, professor (Aizu, Japan), *N. M. Oskorbin*, professor (Barnaul)
D. E. Palchunov, professor (Novosibirsk), *T. Pizansky*, professor (Ljubljana, Slovenia)
V. P. Potapov, professor (Kemerovo), *O. I. Potaturkin*, professor (Novosibirsk)
V. A. Serebryakov, professor (Moscow), *A. V. Starchenko*, professor (Tomsk)
S. I. Smagin, professor, corresponding member of RAS (Khabarovsk)
D. A. Tusupov, professor (Astana, Kazakhstan)
V. V. Shajdurov, professor, corresponding member of RAS (Krasnoyarsk)
Yu. I. Shokin, professor, academician (Novosibirsk)

*The journal is published quarterly in Russian since 1999
by Novosibirsk State University Press*

*The address for correspondence
Faculty of Information Technologies, Novosibirsk State University
1 Pirogov Street, Novosibirsk, 630090, Russia
Tel. +7 (383) 363 42 46*

E-mail address: inftech@vestnik.nsu.ru

On-line version: <http://elibrary.ru>

Научная статья

УДК 004.052.42

DOI 10.25205/1818-7900-2024-22-1-5-20

Построение комплекса автоматизированной отладки фрагментированных программ

Андрей Юрьевич Власенко¹, Михаил Антонович Мичуров²
Василий Дмитриевич Царёв³, Максим Андреевич Курбатов⁴

¹⁻⁴ Новосибирский государственный университет

¹ Институт вычислительной математики и математической геофизики СО РАН
Новосибирск, Россия

¹ a.vlasenko@g.nsu.ru, <https://orcid.org/0000-0001-8247-5736>

² m.michurov@g.nsu.ru, <https://orcid.org/0000-0002-0404-6802>

³ v.tsarev@g.nsu.ru

⁴ m.kurbatov@g.nsu.ru

Аннотация

В статье кратко изложена концепция фрагментированного программирования, а также принципиальное устройство системы автоматического конструирования параллельных программ LuNA (Language for Numerical Algorithms) и одноименного языка программирования. Описаны такие свойства системы LuNA, как возможность работы на вычислительных системах разных типов, динамическая балансировка нагрузки на узлы и процессорные ядра вычислительного кластера и другие.

Целью описываемой в статье работы является создание программного комплекса отладки фрагментированных программ в системе LuNA. В связи с этим приведен ряд ошибок, специфических для фрагментированных программ. Перечислены подходы к отладке параллельных программ и проанализирована их применимость к фрагментированным программам.

Подробно изложена реализация методов статического анализа и «посмертного анализа» в авторских средствах автоматизированной отладки фрагментированных программ для системы LuNA. Два средства статического анализа используют такие промежуточные представления, как абстрактное синтаксическое дерево и граф информационных зависимостей. Третье средство базируется на оригинальном методе статического анализа, заключающемся в генерации Prolog-программы, содержащей набор фактов об исходной LuNA-программе и соответствующие ошибочным ситуациям правила. При последующем запуске Prolog-программы факты проверяются на соответствие правилам, в результате чего пользователю выдаются сообщения о найденных ошибках. Представлена модель интеграции созданных средств в единый комплекс автоматизированной отладки фрагментированных программ, а также утилита автоматизированного тестирования инструментов отладки LuNA-программ.

Ключевые слова

фрагментированное программирование, система LuNA, логические ошибки, автоматизированная отладка, статический анализ, «посмертный анализ».

Финансирование

Исследования выполнены в рамках государственного задания ИВМиМГ СО РАН 0251-2022-0005.

Для цитирования

Власенко А. Ю., Мичуров М. А., Царёв В. Д., Курбатов М. А. Построение комплекса автоматизированной отладки фрагментированных программ // Вестник НГУ. Серия: Информационные технологии. 2024. Т. 22, № 1. С. 5–20. DOI 10.25205/1818-7900-2024-22-1-5-20

© Власенко А. Ю., Мичуров М. А., Царёв В. Д., Курбатов М. А., 2024

Constructing a Complex for Automated Debugging of Fragmented Programs

Andrey Yu. Vlasenko¹, Mihail A. Michurov²
Vasiliy D. Tsarev³, Maksim A. Kurbatov⁴

¹⁻⁴Novosibirsk State University

¹Institute of Computational Mathematics and Mathematical Geophysics SB RAS
Novosibirsk, Russian Federation

¹a.vlasenko@g.nsu.ru, <https://orcid.org/0000-0001-8247-5736>

²m.michurov@g.nsu.ru, <https://orcid.org/0000-0002-0404-6802>

³v.tsarev@g.nsu.ru

⁴m.kurbatov@g.nsu.ru

Abstract

The paper is devoted to the authors' tools for automated debugging of fragmented programs in the LuNA (Language for Numerical Algorithms) system. The LuNA system uses approach of fragmented programming, which allows the researcher to design parallel programs in an automated mode.

The main problem discussed in the paper is the detection of errors specific to LuNA programs. The operation algorithms of four developed tools for automated debugging of LuNA programs are described. Three of them are based on the static analysis approach and differ in the intermediate representations used. The first uses an abstract syntax tree (AST). The second uses a data dependency graph (DDG). The third tool generates a Prolog program containing a set of facts about the original LuNA program and rules corresponding to erroneous situations. Due to the use of various intermediate representations, each of the static analysis tools is able to detect its own types of errors. The last tool uses the "post-mortem analysis" approach. Each process of a parallel program collects a trace file during its operation and at the end of the work these files are analyzed by a special utility.

The paper also describes the developed utility for testing automated debugging tools. The set of test programs with various errors, as well as error-free programs, has been developed for this utility.

The last part describes the designed mechanism for integrating the developed automated debugging tools into a single software package.

Keywords

fragmented programming, logical errors, automated debugging, static analysis, "post-mortem analysis"

Funding

This work was carried out under state contract with ICMMG SB RAS 0251-2022-0005.

For citation

Vlasenko A. Y., Michurov M. A., Tsarev V. D., Kurbatov M. A. Constructing a complex for automated debugging of fragmented programs. *Vestnik NSU. Series: Information Technologies*, 2024, vol. 22, no. 1, pp. 5–20 (in Russ.) DOI 10.25205/1818-7900-2024-22-1-5-20

Введение

Технология фрагментированного программирования подразумевает разделение программы на части, называемые фрагментами кода (ФК). Во время работы программы каждый запущенный экземпляр ФК представляет собой самостоятельную динамическую единицу, называемую фрагментом вычислений (ФВ или CF – calculation fragment). При этом если по коду программы ФК1 предшествовал ФК2, то из этого не следует, что соответствующий ФВ2 будет выполнен строго после ФВ1. Если между ними нет зависимостей по данным, то эти ФВ могут выполняться параллельно, и ФВ2 может завершиться даже раньше, чем ФВ1. ФВ могут быть запущены на различных узлах мультимпьютера, ядрах мультипроцессора или графических ускорителях.

Еще одной важной сущностью во фрагментированной программе является фрагмент данных (ФД или DF – data fragment), представляющий собой некий аналог переменной в языках императивного программирования. Особенность ФД состоит в том, что это объект единствен-

ного присваивания. Любой ФВ потребляет на вход некоторое множество ФД (в том числе и пустое) и выдает на выходе другие ФД, которые, в свою очередь, могут быть потреблены следующими ФВ. Таким образом, ФВ_n является зависимым по данным от ФВ_m, если среди входных аргументов ФВ_n есть ФД, которые вычисляет ФВ_m, либо существует последовательность ФВ₁, ФВ₂, ..., ФВ_k таких, что ФВ₁ зависим по данным от ФВ_m, ФВ₂ – от ФВ₁, ..., ФВ_n – от ФВ_k. ФД, как и ФВ, должны обладать возможностью миграции между узлами мультимпьютера в целях перераспределения вычислительной нагрузки.

Система LuNA и одноименный язык, разрабатываемые в ИВМиМГ СО РАН и на кафедре параллельных вычислений ФИТ НГУ, реализуют технологию фрагментированного программирования [1; 2]. При помощи языка LuNA описывается рекурсивно-перечислимое множество триплетов $\langle in, mod, out \rangle$, где *in* – множество входных ФД, *out* – множество выходных, а *mod* – модуль, реализующий ФК, который вычисляет *out*, если ему на вход поступило *in*. Если множество *in* не полное (не все входные ФД уже вычислены), то *mod* просто ожидает готовности недостающих ФД. ФД могут иметь тип целое число, число с плавающей точкой и строковый тип.

В описываемой системе ФК могут быть как структурированными (на языке LuNA), так и атомарными (на C/C++ или Fortran). Для ФК определены входные и выходные переменные, являющиеся его формальными параметрами. Также ФК не должен иметь «побочных эффектов», т. е. не должен изменять ничего (глобальных переменных, переменных среды, системных настроек и т. д.), кроме своих выходных параметров. Данное требование, как и требование единственности присваивания значений ФД, необходимо для обеспечения возможности исполнения порождаемых ФВ в произвольном порядке, в том числе и параллельно.

Фрагментированная программа в системе LuNA может работать на параллельной вычислительной системе за счет перераспределения ФВ по узлам вычислительного кластера, процессорным ядрам одного узла и графическим ускорителям. При этом динамическим распределением ФВ по вычислительным устройствам занимается runtime-система. В этом заключается балансировка нагрузки.

Листинг 1. Пример ошибочной LuNA-программы

```
1: import c_print(real) as print; // импорт атомарных ФК на C/C++
2: import c_init(real, name) as init; // инициализация ФД
  (второго аргумента) вещественным числом
3: import c_bar(string) as bar;

4: sub main(int a){
5:     df a, b, c, d, e; // объявление ФД
6:     foo(a);
7:     bar(1);
8:     init(1, b);
9:     init(2, b);
10:    print(e);
11: }

12: sub foo(name x){
13:     init(1, x);
14:     print(x);
15: }
```

Создание LuNA-программ сопряжено с риском написания ошибочного кода. Ввиду особенностей языка LuNA-программам свойственны специфические семантические ошибки, не характерные для стандартных языков и технологий параллельного программирования.

Авторами статьи была создана база данных с ошибками, их описанием и исходным кодом, их воспроизводящим. На текущий момент база данных содержит 22 различные ошибки, из которых 4 будут рассмотрены ниже.

Демонстрация алгоритмов и структур будет произведена на основе программы из листинга 1, содержащей следующие ошибки.

1. Попытка использования неинициализированного ФД. Эта ошибка приводит к зависанию программы, поскольку ФВ, ожидающий получения неинициализированного ФД, вынужден ожидать его бесконечно. Пример: строка 10 – ФД *e*.

2. Повторная инициализация ФД. В случае инициализации уже инициализированного ФД программа может завершиться преждевременно. Такую ситуацию пользователю может быть тяжело обнаружить визуально в сложной программе с большим количеством ветвлений. Пример: строка 9 – ФД *b* уже был инициализирован в строке 8.

3. Неиспользуемый ФД. Данная ситуация не является ошибочной в полном смысле слова, но тем не менее является примером неоптимального поведения и, что важнее, может указывать на наличие иных ошибок, связанных с указанным ФД. Пример: строка 5 – ФД *d*.

4. Несоответствие типов аргументов при вызове атомарного ФК. Возникает в случае передачи в ФК аргумента иного типа, нежели ожидался. Пример: строка 7 – несоответствие константной единицы и типа *string*.

Фрагментированное программирование – новый перспективный подход для построения программ, способных эффективно использовать высокопроизводительные вычислительные ресурсы. А в связи с тем что программам в этой технологии свойственны специфические ошибки, острым вопросом встает создание средств отладки. Авторы данной статьи задались целью разработки комплекса средств автоматизированного обнаружения ошибок в LuNA-программах. Автоматизированная отладка существенно ускорит процесс разработки как для новых, так и для уже опытных пользователей системы LuNA. Программный комплекс автоматизированной отладки будет построен на основе интеграции четырех созданных инструментальных средств (AST-analyzer, DeGSA, Prolog-analyzer и luna-trace), каждое из которых будет описано ниже.

1. Методы отладки

Для обнаружения ошибок в программах существуют различные инструментальные средства, которые можно классифицировать по нескольким используемым методам. С целью выбора подходов, пригодных к использованию при построении средств автоматизированной отладки, был выполнен нижеследующий обзор и анализ.

Самым распространенным является метод диалоговой отладки, который используют такие популярные системы, как TotalView и Linaro Distributed Debugging Tool (DDT)¹. При работе с подобной системой пользователь расставляет в программе контрольные точки, движется по шагам, анализируя изменения значений переменных и др. Такой анализ хорошо знаком большинству программистов, соответственно системы диалоговой отладки обладают низким входным порогом использования. Однако этот подход плохо применим в условиях больших суперкомпьютерных центров, поскольку там пользователи ставят свои задачи в очередь и на этом теряют над ними контроль. Возможности интерактивного взаимодействия с исполняющейся программой при этом сильно ограничены. Кроме того, на сложных параллельных программах встречаются такие ситуации, когда на малых размерностях при запуске небольшого количества процессов программа отработывает корректно, а при переходе на сотни процессов проявляется ошибка. Методом диалоговой отладки выявлять подобные ошибки крайне трудоемко.

¹ Get started with DDT. URL: https://docs.linaroforge.com/23.0.2/html/forge/ddt/get_started_ddt/index.html

В связи с этим перспективными представляются автоматизированные методы отладки, не требующие непосредственного диалога пользователя с работающей программой. Некоторые из них будут изложены ниже.

Статический анализ выполняется над исходным кодом или над одним из промежуточных представлений программы (абстрактным синтаксическим деревом – AST [3], трехадресным кодом, графом потока управления и др.). Таким образом, статический анализ производится до непосредственного запуска программы.

Глубина статического анализа может варьироваться от определения поведения отдельных операторов до анализа всего имеющегося исходного кода. В коммерческой среде статический анализ главным образом применяется для поиска уязвимостей в системах, критичных к безопасности, однако при помощи данного подхода возможно обнаруживать различные семантические ошибки. Примерами статических анализаторов являются Clang-Tidy, PVS-Studio, PC-LINT и др.

Очевидными преимуществами данного метода являются:

- возможность полного покрытия кода – при динамическом анализе программа проходит только один из всех возможных вариантов;
- проверка на ошибки без запуска анализируемого приложения (параллельные программы для суперкомпьютерных систем очень часто работают весьма значительное время);
- независимость от конкретного компилятора и среды исполнения.

По этим причинам три разрабатываемых и описанных ниже средства обнаружения ошибок в LuNA-программах используют подход статического анализа. Принципиальное их различие – в промежуточных представлениях, на базе которых осуществляется анализ.

К минусам же метода статического анализа относятся так называемые «ложные срабатывания», когда средство выдает предупреждения на фактически нормальное поведение.

Оригинальным и перспективным вариантом статического анализа является генерация по исходному коду в одном языке программы на каком-либо ином языке (как правило – в другой парадигме), после чего производится стандартная компиляция с последующим запуском или интерпретация сгенерированной программы. В процессе работы обнаруживаются заданные правилами ошибочные ситуации [4]. Такой подход, в частности, применяется в еще одном инструментальном средстве отладки фрагментированных программ, генерирующем Prolog-программу и описанном далее в п. 5.

Еще одним подходом к автоматизированному обнаружению ошибок в программах (в том числе и параллельных) является сравнительная отладка. Суть этого метода заключается в том, что инструментальному средству дается 2 версии программы: эталонная и отлаживаемая. Средство сравнивает значения переменных обеих программ в определенных пользователем контрольных точках и сигнализирует при несовпадении. В том числе возможна и такая ситуация, что за эталонную будет взята последовательная программа, а за отлаживаемую – параллельная. Сравнению чаще всего подвергаются трассы обеих программ с записанными значениями переменных в контрольных точках. Характерными представителями данного подхода являются отладчик Guard [5] и система DVM [6]. Данный метод на текущем этапе не рассматривается в качестве возможного для внедрения в разрабатываемый комплекс автоматизированной отладки LuNA-программ в связи с тем, что эталонная версия, как правило, отсутствует.

Следующий метод – автоматизированный (или динамический) контроль корректности [7], при котором анализируемая программа проверяется инструментальным средством на соответствие определенным правилам, заложенным в инструментальном средстве. Примерами таких средств являются Intel Trace Analyzer and Collector (ITAC) и MUST [8], проверяющие MPI-программу на соблюдение ограничений MPI-стандарта. Данный анализ можно выполнять по ходу работы программы (в online-режиме) или по собранной трассе после завершения программы («посмертный анализ»). Поскольку ряд семантических ошибок крайне сложно обнаружить методом статического анализа, то одно из разрабатываемых инструментальных средств отладки

использует именно подход автоматизированного контроля корректности (см. п. 3), благодаря чему расширяется охват возможных проблем в LuNA-программах.

Последним рассмотренным в данной статье методом будет верификация программ на моделях (Model Checking) [9]. При использовании данного подхода формулируются требования к анализируемой системе (в частном случае – программе), для чего зачастую применяется аппарат темпоральной логики. В то же время строится модель на некотором псевдоязыке, включающая только те свойства системы, которые могут повлиять на соблюдение требований в процессе функционирования системы. Данная модель обладает конечным числом состояний и представима в виде графа. Затем верификатор выполняет исчерпывающий поиск в данном графе тех путей, которые ведут к нарушению требований (так называемых «контр-примеров»). Одной из основных проблем метода верификации программ на моделях является «комбинаторный взрыв» пространства состояний – экспоненциальный рост числа состояний и путей в результирующем графе в зависимости от количества параллельных ветвей исследуемой программы и количества переменных, принимаемых во внимание в модели. Одним из самых популярных верификаторов является Spin [10]. Данный метод в последующем также планируется к внедрению в разрабатываемый комплекс инструментальных средств отладки LuNA-программ.

2. Анализ программы по собранной трассе

Система LuNA состоит из нескольких взаимодействующих модулей и позволяет легко внедрять новые модули, ответственные за какую-либо функциональность. Такой подход к проектированию системы позволил внедрить модуль трассировки WorkflowTracer, записывающий информацию об исполнении программы в файлы трассы. Собранная информация в дальнейшем анализируется утилитой luna_trace [7] с целью обнаружения семантических ошибок, имевших место во время работы программы.

Модуль WorkflowTracer предоставляет интерфейс для уведомления о значимых для последующего анализа событиях, таких как создание и удаление ФВ и ФД, пересылка ФВ между узлами и т. п., и собирает трассу на каждом узле.

По завершении работы программы на каждом вычислительном узле создается три файла трассы со следующей информацией:

- 1) список ФВ, выполненных на данном узле;
- 2) список ФД, созданных на данном узле;
- 3) список ФВ, выполнение которых на данном узле было начато, но не завершилось.

Зависание системы может быть вызвано наличием ошибки использования неинициализированного ФД. В случае же, когда зависание вызвано иными причинами (например, бесконечным циклом в коде атомарного ФВ на C++), пользователь может остановить работу программы, пошлав сигнал SIGINT (например, используя комбинацию «Ctrl» + «C»). В этом случае вся информация, накопленная в буферах отложенной записи для файлов трасс, будет записана на диск.

В исполнительной runtime-системе LuNA за остановку отвечает модуль IdleStopper, реализующий адаптированную версию алгоритма Дейкстры – Шольтена [11]. Алгоритм оперирует понятиями «наличия» и «отсутствия» работы на вычислительных узлах. Когда на всех узлах не остается работы, система может быть остановлена. В системе LuNA наличие работы на каждом узле определяется счетчиком незавершенных ФВ на данном узле. Если ФВ ожидает данные, то он не учитывается счетчиком. Таким образом, модуль IdleStopper будет считать, что на узле не осталось работы, если на нем либо все ФВ завершены, либо остались только те, которые ожидают данные. Когда на всех узлах не осталось работы, то IdleStopper завершает

программу. По наличию ожидавших данные ФВ WorkflowTracer делает вывод о зависании системы.

Утилита анализа трасс, названная `luna_trace`, агрегирует трассы, собранные на узлах мультимпьютера во время работы программы, и производит их анализ для выявления семантических ошибок. Утилита предоставляет информацию не только о факте наличия ошибок как таковых, но, по возможности, и о причинах возникновения ошибок с привязкой к исходным кодам LuNA-программы, а также о возможных способах исправления.

Поиск причин зависаний

Под зависшим ФВ будем понимать ФВ, выполнение которого не может завершиться при текущем состоянии исполнительной системы (например, если нет необходимых ФД). Под причиной зависания понимается зависший ФВ, не имеющий информационных зависимостей от других зависших ФВ.

Для поиска причин зависания строится ориентированный граф информационных зависимостей между зависшими ФВ. Тогда причинами зависаний будут ФВ, соответствующие вершинам, не имеющим исходящих ребер.

В зависимости от опций запуска `luna_trace` может как вывести информацию обо всех зависших ФВ, так и только о причинах. Возможна ситуация, когда в непустом графе нет ни одной вершины, не имеющей исходящих ребер, т. е. причины зависания выделить невозможно. В этом случае `luna_trace` выводит информацию обо всех зависших ФВ, а также сообщение о наличии циклических зависимостей между ФВ.

Потенциальные места инициализации данных

Инструментальное средство `luna_trace` при обнаружении попытки использования неинициализированного ФД выполняет поиск мест в коде программы, где могла быть пропущена инициализация. Этот поиск основан на обнаружении мест инициализации *похожих* ФД.

Пусть существуют множества ФД:

$$DF_1 = \{ \text{name}[i] \mid i \in I \},$$

$$DF_2 = \{ \text{name}[j] \mid j \in J, I \cap J = \emptyset \}.$$

Пусть ФД из множества DF_1 не были инициализированы во время работы программы, а ФД из множества DF_2 были инициализированы множеством ФВ CF , причем все ФВ из множества CF выполняли один и тот же код (например, операцию инициализации i -го индекса ФД в цикле по i). Тогда можно предположить, что ФД из множества DF_1 должны были быть инициализированы фрагментами вычислений, выполняющими тот же код, что и ФВ из множества CF . Собранная во время выполнения программы трасса содержит информацию, по которой можно восстановить потенциальные места инициализации в коде.

Поиск случаев множественной инициализации данных

В файлах трассы для создаваемых во время работы программы ФВ имеется информация об идентификаторах входных и выходных ФД. Наличие этой информации позволяет обнаружить повторяющиеся ФД среди выходных, что соответствует ошибке множественной инициализации.

3. Статический анализ на базе AST

Для применения статического анализа необходимо иметь удобное представление исходного кода. В качестве промежуточного представления можно использовать абстрактное синтаксическое дерево (AST), трехадресный код, граф потока управления (CFG) и др. [12]. Различные представления подходят для обнаружения ошибок различных типов. В связи с этим в создаваемый комплекс автоматизированной отладки будут включены 3 разрабатываемых авторами статических анализатора (AST-analyzer, DeGSA – Dependence Graph Static Analyzer, Prolog-analyzer).

Абстрактное синтаксическое дерево строится по известному алгоритму [13]. Это конечное помеченное ориентированное дерево, в котором внутренние вершины сопоставлены с операторами языка программирования, а листья – с соответствующими операндами.

Другой важной функцией анализа AST является определение структуры программы. Например, можно определить, какие функции присутствуют в программе и какие аргументы они принимают, или выяснить, какие классы определены в коде.

По коду из листинга 1 AST-analyzer строит следующее дерево (не принципиальные элементы опущены) (рис. 1).

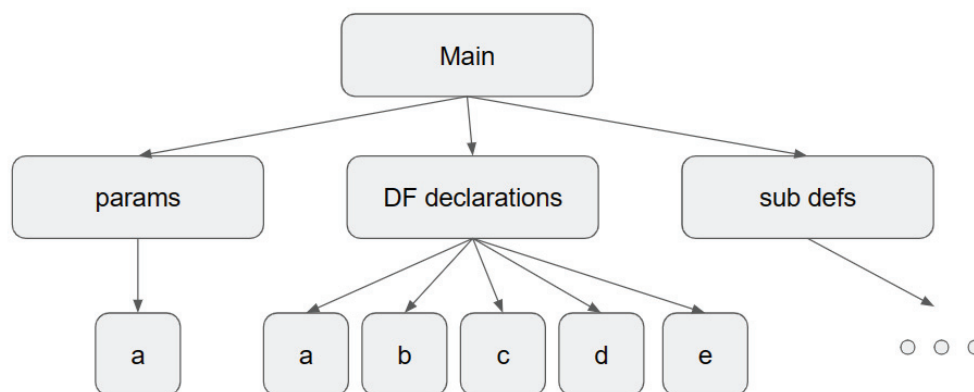


Рис. 1. AST для программы (1)

Fig. 1. AST for program (1)

Здесь корнем дерева является ФК Main, у которого есть параметр *a*, декларации фрагментов данных (5шт.) и вызовы других фрагментов кода.

Далее совершается обход этого AST в глубину с сохранением в контейнер всех параметров и объявлений ФД. После этого линейным поиском выявляются все повторяющиеся декларации.

AST-analyzer способен обнаруживать такие ошибки, как:

- 1) импорт двух функций под одним именем;
- 2) отсутствие функции main;
- 3) ФД с одинаковыми именами в одной области видимости;
- 4) несовпадение количества аргументов при объявлении ФК и его вызове.

Однако не все ошибки можно обнаружить методом статического анализа по AST. Например, это касается ошибки «Несоответствие типов аргументов при вызове структурированного ФК», поскольку не всегда тип ФД можно определить до запуска программы. Таким образом, если *X*, *Y* – ФД и тип хотя бы одного из них не известен, то невозможно определить тип ариф-

метического выражения ($X + Y$, $X * Y$ и т. д.). Возможно, X или Y будут инициализированы строковым типом, и выражение вообще не будет иметь смысл. Поэтому если найдена декларация и вызов фрагмента кода, то при выявлении несовпадения типов параметров анализатор предупредит программиста о возможной ошибке. Другим примером ошибки, для обнаружения которой не подходит такое промежуточное представление, как AST, является попытка использования неинициализированного ФД.

4. Статический анализ на базе графа зависимостей по данным

Одним из вариантов альтернативного промежуточного представления, которое было рассмотрено авторами на предмет возможности применения к системе LuNA, является *граф потока управления* (control-flow graph, CFG) [14]. В традиционных императивных языках программирования (таких как C++ или Java) CFG представляет собой ориентированный граф, вершинами которого являются базовые блоки, а дуги указывают передачу управления от одного базового блока к другому. CFG строится из трехадресного кода, который, в свою очередь, строится из AST.

Однако при построении трехадресного кода LuNA-программы возникают трудности, самой главной из которых является недетерминированность порядка выполнения ФВ. Трехадресный код опирается на конкретный порядок выполнения команд, и, как следствие, CFG тоже этого требует. Это условие делает CFG неприменимым для статического анализа LuNA-программ.

В связи с этим авторами в качестве альтернативы был выбран граф зависимостей по данным [15] (data-dependence graph, DDG), который применяется в различных языках программирования (в частности для оптимизаций) и представляет собой ориентированный граф, где вершинами выступают некоторые операции, а дугами – зависимости операндов операций от результатов вычислений других операций. Такой граф также возможно построить из AST. Авторское инструментальное средство, использующее DDG для обнаружения семантических ошибок в LuNA-программе, носит название DeGSA (Dependence Graph Static Analyzer).

При построении DDG был выбран подход, при котором каждая вершина графа представляет собой оператор, который имеет входные вершины (от которых он зависит) и выходные (которые зависят от него), а также ссылается на все операторы, которые находятся в его теле (далее – «дочерние» вершины).

В качестве примера ниже продемонстрирован DDG для программы из листинга 1 (рис. 2):

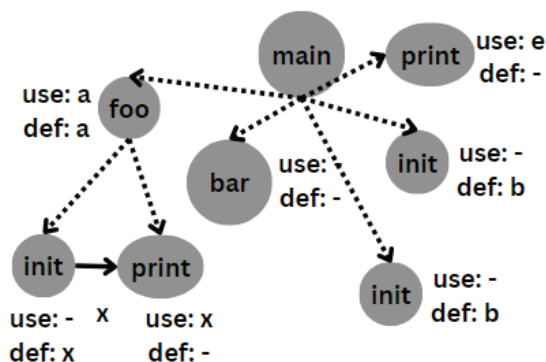


Рис. 2. Граф зависимостей по данным для программы (1)

Fig. 2. Data-dependence graph for program (1)

В этом графе сплошными стрелками указаны дуги зависимостей по данным, а пунктирными стрелками – нахождение одной вершины внутри блока второй (например, *foo* и *bar* находятся внутри блока *main*).

Ниже рассмотрен алгоритм построения и применения DDG в средстве DeGSA, разделяющийся на четыре этапа.

1. Линейный обход всех атомарных ФК и анализ их сигнатур. А именно поиск и сохранение информации о том, какие аргументы при вызове атомарного ФК будут использованы, а какие инициализированы.

2. Создание вершин. Вершина создается только после того, как все операторы в теле текущего оператора были также пройдены, и для них были созданы вершины. Процесс создания различается в зависимости от типа оператора. Для примера далее рассмотрен оператор *cf*, т. е. оператор применения ФК – порождения ФВ. Если ФК является атомарным, то вершина создается сразу. По сигнатуре атомарного ФК можно однозначно определить, является ли каждый из аргументов инициализируемым (имеет тип *name*) или используемым (любой другой тип). Если ФК является структурированным, то функция рекурсивно запускает себя для анализа операторов в теле текущего и порождает дочерние вершины, после чего создает и вершину текущего оператора. Для каждой вершины сохраняется информация, какие именно ФД были использованы и инициализированы в соответствующем ей операторе. Также на этом этапе сохраняется информация о вершинах, содержащихся в блоке текущей вершины (эта связь и указывается на рис. 2 пунктирными стрелками).

3. Рекурсивный обход всех созданных вершин и их связывание. Дуга (сплошная стрелка) от первой вершины до второй создается в случае, если у первой вершины ФД инициализируется, а у второй он же используется (в примере на рисунке в *init* инициализируется ФД *x*, а в *print* – используется).

4. Поиск ошибок.

Для демонстрации возможностей полученного графа ниже рассмотрен поиск двух видов возможных ошибок на примере программы из листинга 1.

1. Повторная инициализация ФД. DeGSA обнаруживает подобные ситуации путем поиска по графу в глубину всех вершин, где ФД указан как используемый. На рис. 2 можно заметить, что от вершины *main* идут две дуги к вершинам *init*, каждая из которых инициализирует ФД *b*, что является ошибкой.

2. Неиспользуемый ФД. Здесь DeGSA ищет вершины, из которых нет исходящей дуги с меткой некоторого инициализируемого ФД. Анализ графа позволяет заметить, что из двух вершин, инициализирующих ФД *b*, не идет никаких дуг, что позволяет заключить, что фрагмент данных *b* не используется и может быть удален.

5. Статический анализ с использованием Prolog

В рамках еще одного разрабатываемого инструмента статического анализа LuNA-программ Prolog-analyzer было решено использовать язык логического программирования Prolog. Использование Prolog позволяет формализовать семантические ошибки в виде лаконичных правил. Еще одним плюсом данного подхода является то, что при проверке этих правил Prolog инстанцирует входящие в правила свободные переменные. Это позволяет получить как информацию о наличии ошибок, так и информацию о том, для каких значений входных ФД найденные ошибки имеют место.

Формат представления LuNA-программы

Существуют работы, в которых языки логического программирования (в частности, Prolog и Datalog) используются для статического анализа [16]. В данной работе статический анализа-

тор по AST LuNA-программы строит набор фактов (часть базы знаний), описывающих инициализацию и использование ФД.

В качестве анализируемых абстракций были выбраны циклы инициализации и использования, а также вызовы атомарных ФК, инициализирующих и использующих одиночные ФД (не содержащие индексов). Для представления LuNA-программы используются следующие факты:

- *df_init_single*(InitID, Name) – инициализация одиночного ФД;
- *df_use_single*(UseID, Name) – использование одиночного ФД;
- *df_init_range*(InitId, Name, boundary(InitStart, InitStartType), boundary(InitEnd, InitEndType)) – инициализация диапазона индексов (цикл инициализации) для ФД Name, начиная от InitStart и заканчивая InitEnd;
- *df_use_range*(UseId, Name, boundary(UseStart, UseStartType), boundary(UseEnd, UseEndType)) – использование диапазона индексов (цикл использования) для ФД Name, начиная от UseStart и заканчивая UseEnd.

Во всех четырех видах фактов *InitID* и *UseId* – уникальные идентификаторы мест инициализации и использования ФД в программе, а *Name* – имя ФД. Параметры {Init|Use} {Start|End} Type могут принимать значения «const», «df» и «expr».

В данный момент реализована генерация базы знаний для ограниченного подмножества языка LuNA: поддерживаются вызовы ФК (как атомарных, так и структурированных) с параметрами типа *name* и параметрами примитивных типов, и оператор *for*.

Ниже рассматривается ошибка использования неинициализированных ФД. Правила для поиска двух ее разновидностей, записываемые в код на Prolog:

- *use_of_uninitialized_df*(Name, Details) – использование неинициализированного одиночного ФД;
- *invalid_df_use_range*(Name, Details) – некорректные границы цикла использования ФД.

Name – имя ФД, *Details* – подробная информация об ошибке, например, информация о том, верхняя или нижняя граница цикла использования была некорректной, а также конкретные выражения, использованные в качестве границы.

Выражения, используемые в качестве границ, сравниваются с помощью библиотеки CLP(FD) (Constraint Logic Programming over Finite Domains). Различаются ситуации, когда границы циклов инициализации и использования можно сравнить однозначно (например, N и $N + 1$) и когда это невозможно без знания конкретных значений выражений (например, 100 и N).

Автоматизация генерации базы знаний и выполнения запросов

Инструментальное средство Prolog-analyzer автоматизирует процесс генерации Prolog-представления LuNA-программы и проверки выполнимости правил для поиска ошибок. Оно реализовано на языке Python как утилита командной строки.

Средство анализа генерирует и сохраняет информацию об использовании ФД (места использования/инициализации). Затем генерируется база знаний на языке Prolog. После к базе знаний выполняются запросы (в частности проверяется выполнимость правил *invalid_df_use_range* и *use_of_uninitialized_df*), и выводится подробная информация об обнаруженных ошибках.

6. Утилита автоматизированного тестирования инструментальных средств отладки LuNA-программ

Для тестирования создаваемых средств отладки LuNA-программ на корректность работы была создана утилита автоматизированного тестирования. Утилита получает на вход директо-

рию, содержащую файл конфигурации *test_config.json*. Конфигурация теста состоит из четырех основных частей:

- 1) набор команд, которые нужно выполнить перед тестом (поле «arrange»);
- 2) команда, работа которой тестируется (поле «act»);
- 3) проверки, которые необходимо выполнить над выводом тестируемой команды (поле «assert»);
- 4) список имен потоков вывода («stderr», «stdout»), которые нужно сохранить в файлы на диске (поле «save»).

В листинге 2 приводится пример конфигурации теста. В данном случае требуется, чтобы вывод анализатора был пустым, так как в LuNA-программе нет ошибок, которые он мог бы обнаружить.

Листинг 2. Конфигурация теста для программы, не содержащей ошибок

```
0: {
1:   "arrange": [
2:     {
3:       "platform": "Linux",
4:       "command": "./build.sh"
5:     }
6:   ],
7:   "act": {
8:     "command": "prolog-analyzer --project-dir=. --build-dir=bin",
9:     "timeout": 10
10:  },
11:  "assert": {
12:    "stderr": {"empty": true},
13:    "stdout": {"empty": true}
14:  }
15: }
```

Проверки могут производиться над потоками вывода (stdout) и ошибок (stderr). В качестве проверки можно указать регулярное выражение, которое должно содержаться в выводе; путь к файлу шаблона, содержимое которого должно совпадать с выводом команды; факт того, что в поток не должно ничего выводиться, как в листинге 2.

Для проверки корректности разрабатываемых анализаторов был создан набор тестов, каждый из которых содержит LuNA-программу, скрипт для ее компиляции и файл *test_config.json*. В созданный набор тестов вошли LuNA-программы, содержащие все 22 выделенные ошибки, а также безошибочные программы.

После того как вывод всех средств отладки будет осуществляться в едином JSON-формате (см. п. 7), шаблоны проверок в тестах будут переписаны и утилита автоматизированного тестирования сможет унифицированным образом работать с любым из описанных средств отладки.

7. Интеграция утилит автоматизированной отладки в программный комплекс

Прорабатывается интеграция описанных выше анализаторов в единый комплекс автоматизированной отладки LuNA-программ, способный применять различные подходы к отладке и агрегировать результаты их работы. В целях данной интеграции ведется работа по унификации вывода разработанных анализаторов в JSON-формате, что позволит объединять вывод

и предоставлять пользователю максимально полный и информативный отчет обо всех ошибках в LuNA-программе.

Анализаторы также будут взаимодействовать друг с другом, обмениваясь вспомогательной информацией о найденных и/или возможных ошибках. Примером такого взаимодействия является проверка значений конкретных выражений во время работы программы в случае, когда от конкретного значения будет зависеть наличие или отсутствие ошибочного поведения. Например, Prolog-analyzer устанавливает возможную ошибку несоответствия границ циклов использования и инициализации в зависимости от значения некоторого параметра N . luna_trace, в свою очередь, будет проверять факт наличия этой ошибки, поскольку может получить информацию о реальном значении, которое N принимало в программе.

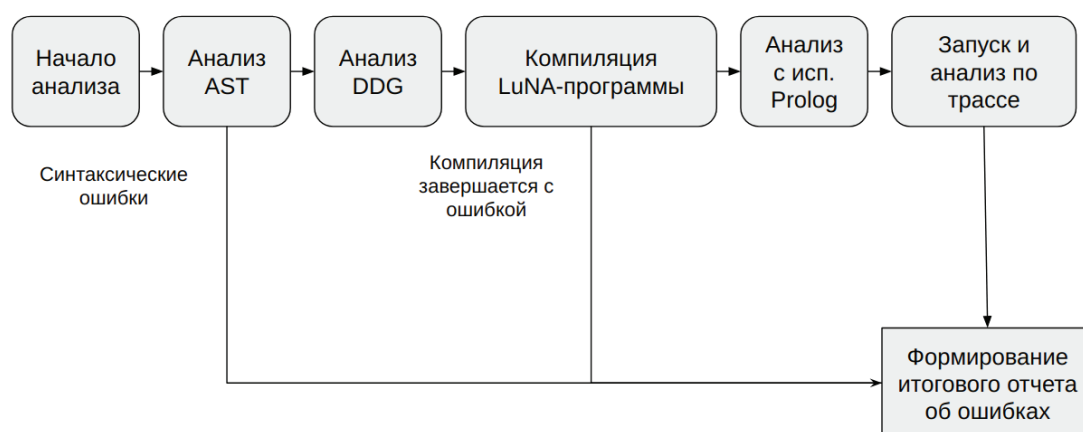


Рис. 3. Схема работы комплекса автоматизированной отладки

Fig. 3. The operation scheme of the automated debugging complex

Схема работы программного комплекса автоматизированной отладки изображена на рис. 3. Сначала запускается AST-analyzer. На данном этапе будут обнаружены такие ошибки, как отсутствие объявления ФК *main*, импортирование нескольких атомарных ФК под одним именем, и другие. Информация об обнаруженных ошибках будет записываться в JSON-файл отчета. Затем запускается DeGSA. На этом этапе будут обнаруживаться более сложные ошибки, такие как повторная инициализация ФД, а также неиспользуемые ФД. Соответственно будет дополнен JSON-файл отчета. Далее LuNA-программа проходит стадии парсинга и препроцессинга с последующей генерацией ее внутреннего представления, используемого компилятором. Данное представление также используется для генерации Prolog-программы третьим статическим анализатором. Автоматизированное выполнение запросов к составленной на следующем шаге базе знаний на Prolog позволит обнаружить еще ряд ошибок, которые дополняют JSON-отчет. Наконец, LuNA-программа будет скомпилирована и запущена, после чего luna_trace произведет анализ трассы и завершит заполнение файла отчета. Затем JSON-файл будет потреблен программным средством, формирующим человекочитаемый отчет о найденных ошибках и являющийся конечным результатом работы создаваемого программного комплекса.

Заключение

Инструментальные средства разрабатываемого программного комплекса автоматизированной отладки LuNA-программ на данный момент способны обнаруживать ряд часто встречающихся ошибок. Однако авторами ведутся работы по расширению функциональных

возможностей данных средств и, в первую очередь, в отношении поиска других ошибок, характерных для LuNA-программ. Благодаря тому что разрабатываемые анализаторы следуют подходу автоматизированной отладки, их становится возможным интегрировать в рамках единого комплекса. А благодаря тому что они используют различные методы (статического и «посмертного» анализа) или различные представления исходного кода, каждый из них подходит для обнаружения своего множества ошибок. Таким образом анализаторы будут дополнять друг друга, формируя единый JSON-файл найденных ошибок. Данный файл будет подаваться на вход средству, формирующему человекочитаемый отчет об ошибках для пользователя.

Для проверки сохранения контроля корректности анализаторов разработана утилита тестирования, включающая тесты на все известные авторам ошибки в LuNA-программах. Благодаря тому что все анализаторы будут выдавать информацию об ошибках в едином формате, утилита сможет однообразно проверять все разрабатываемые инструментальные средства отладки.

В ближайшие планы авторов входит унификация вывода разрабатываемых средств в JSON-формате и создание средства автоматической генерации подробного отчета об ошибках по JSON-файлу. Также программный комплекс будет протестирован на «больших» LuNA-программах численного моделирования для оценки величины накладных расходов.

Список литературы

1. **Malyshkin V., Perepelkin V., Lyamin A.** Trace Balancing Technique for Trace Playback in LuNA System // Malyshkin V. (Eds) Parallel Computing Technologies. PaCT 2023. Lecture Notes in Computer Science. 2023. Vol. 14098. Springer, Cham. P. 4250. DOI: 10.1007/978-3-031-41673-6_4
2. **Малышкин В. Э.** Технология фрагментированного программирования // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2012. № 46 (305). С. 4555. DOI: 10.14529/cmse120104
3. **Зубов М. В., Пустыгин А. Н., Старцев Е. В.** Применение универсальных промежуточных представлений для статического анализа исходного программного кода // Доклады Томского гос. ун-та систем управления и радиоэлектроники. 2013. № 1 (27). С. 64–68.
4. **Flederer F., Ostermayer L., Seipel D., Montenegro S.** Source Code Verification for Embedded Systems using Prolog // Electronic Proceedings in Theoretical Computer Science. 2016. Vol. 234. P. 88–103. DOI: 10.4204/EPTCS.234.7
5. **Abramson D. A., Watson G., Le P. D.** Guard: a tool for migrating scientific applications to the .NET Framework // Sloot P. M. A., Tan C. J. K., Dongarra J. J., & Hoekstra A. G. (Eds.) Computational Science – ICCS 2002: International Conference Amsterdam, The Netherlands, April 21–24, 2002 Proceedings. P. 834843. (Lecture Notes in Computer Science; Vol. 2330). Springer. DOI: 10.1007/3-540-46080-2_88
6. **Bakhtin V. A., Kataev N. A., Kolganov A. S., Yakobovskiy M. V., Zaharov D. A.** Automation of programming for promising high-performance computing systems // Malyshkin V. (eds) Parallel Computing Technologies. PaCT 2023. Lecture Notes in Computer Science. Cham: Springer, 2023. Vol. 14098. P. 3–17.
7. **Malyshkin V., Vlasenko A., Michurov M.** Automated Debugging of Fragmented Programs in LuNA System // Proc. of Mathematical Modeling and Supercomputer Technologies 2022. Communications in Computer and Information Science. 2022. Vol. 1750. P. 266–280. DOI: 10.1007/978-3-031-24145-1_22
8. **Protze J., Hilbrich T., Schulz M., de Supinski B. R.** MPI Runtime Error Detection with MUST: A Scalable and Crash-Safe Approach // Proc. of 43rd International Conference on

- Parallel Processing Workshops. Minneapolis, MN, USA, 2014, p. 206–215. DOI: 10.1109/ICPPW.2014.37
9. **Карпов Ю. Г.** MODEL CHECKING. Верификация параллельных и распределенных программных систем. СПб.: БХВ-Петербург, 2010. 560 с.
 10. **Holzmann G. J., Joshi R.** Model-Driven Software Verification // Graf S., Mounier L. (Eds) Model Checking Software. Proc. of SPIN 2004. Lecture Notes in Computer Science. Vol. 2989. Springer, Berlin, Heidelberg. P. 76–91. DOI: 10.1007/978-3-540-24732-6_6
 11. **Fokkink W.** Distributed Algorithms, second edition: An Intuitive Approach. MIT Press, 2018. 272 p.
 12. **Касьянов В. Н., Евстигнеев В. А.** Графы в программировании: обработка, визуализация и применение. СПб.: БХВ-Петербург, 2003. 1104 с.
 13. **Белеванцев А. А.** Многоуровневый статический анализ исходного кода программ для обеспечения качества программ // Программирование. 2017. Т. 43, № 6. С. 3–25.
 14. **Krinke J.** Information Flow Control and Taint Analysis with Dependence Graphs. Vancouver: Proc. of 3d International Workshop on Code Based Software Security Assessments (CoBaSSA). 2007. P. 6–9.
 15. **Kuck D. J., Kuhn R. H., Padua D. A., Leasure B., Wolfe M.** Dependence graphs and compiler optimizations // Proc. of the 8th ACM SIGPLAN-SIGACT symposium on Principles of Programming Languages (POPL). 1981. P. 207–218. DOI: 10.1145/567532.567555
 16. **Smaragdakis Y., Bravenboer M.** Using Datalog for Fast and Easy Program Analysis // de Moor O., Gottlob G., Furche T., Sellers A. (Eds) Datalog Reloaded. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2010. Vol. 6702. P. 245–251. DOI: 10.1007/978-3-642-24206-9_14

References

1. **Malyshkin V., Perepelkin V., Lyamin A.** Trace Balancing Technique for Trace Playback in LuNA System. In: *Malyshkin, V. (eds) Parallel Computing Technologies. PaCT 2023. Lecture Notes in Computer Science*, 2023, vol. 14098. Springer, Cham., pp. 42–50. DOI: 10.1007/978-3-031-41673-6_4.
2. **Malyshkin V. E.** Technology of fragmented programming. *Vestnik YuUrGU. Series: Computational Mathematics and Software Engineering*, 2012, no. 46 (305), pp. 45–55 (in Russ.). DOI: 10.14529/cmse120104
3. **Zubov M. V., Pustygyn A. N., Starcev E. V.** The use of universal intermediate representations for static analysis of the source code. *Reports of Tomsk State University of Control Systems and Radioelectronics*, 2013, no. 1(27), pp. 64–68.
4. **Flederer F., Ostermayer L., Seipel D., Montenegro S.** Source Code Verification for Embedded Systems using Prolog. *Electronic Proceedings in Theoretical Computer Science*, 2016, vol. 234, pp. 88–103. DOI: 10.4204/EPTCS.234.7
5. **Abramson D. A., Watson G., Le P. D.** Guard: a tool for migrating scientific applications to the .NET Framework. In *P. M. A. Sloot, C. J. K. Tan, J. J. Dongarra, & A. G. Hoekstra (Eds.) Computational Science – ICCS 2002: International Conference Amsterdam, The Netherlands, April 21–24, 2002 Proceedings*, pp. 834–843. (Lecture Notes in Computer Science; Vol. 2330). Springer. DOI: 10.1007/3-540-46080-2_88
6. **Bakhtin V. A., Kataev N. A., Kolganov A. S., Yakobovskiy M. V., Zaharov D. A.** Automation of programming for promising high-performance computing systems. In: *Malyshkin V. (eds) Parallel Computing Technologies. PaCT 2023. Lecture Notes in Computer Science*, Cham: Springer, 2023, vol. 14098, pp. 3–17.
7. **Malyshkin V., Vlasenko A., Michurov M.** Automated Debugging of Fragmented Programs in LuNA System. *Proc. of Mathematical Modeling and Supercomputer Technologies 2022*.

- Communications in Computer and Information Science*, 2022, vol. 1750, pp. 266–280. DOI: 10.1007/978-3-031-24145-1_22
8. **Protze J., Hilbrich T., Schulz M., de Supinski B. R.** MPI Runtime Error Detection with MUST: A Scalable and Crash-Safe Approach. *Proc. of 43rd International Conference on Parallel Processing Workshops*. Minneapolis, MN, USA, 2014, p. 206–215. DOI: 10.1109/ICPPW.2014.37
 9. **Karpov J. G.** MODEL CHECKING. Verification of parallel and distributed software systems. Saint Petersburg, BHV-Peterburg publ., 2010, 560 p. (in Russ.)
 10. **Holzmann G. J., Joshi R.** Model-Driven Software Verification. In: *Graf S., Mounier, L. (eds) Model Checking Software. Proc. of SPIN 2004. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2004, vol. 2989, p. 7691. DOI: 10.1007/978-3-540-24732-6_6
 11. **Fokink W.** Distributed Algorithms, second edition: An Intuitive Approach. MIT Press, 2018, 272 p.
 12. **Kas'janov V. N., Evstigneev V. A.** Graphs in programming: processing, visualization and application. Saint Petersburg, BHV-Peterburg publ., 2003, 1104 p. (in Russ.)
 13. **Belevancev A. A.** Multilevel static analysis of the source code to ensure the quality of programs. *Programming and Computer Software*, 2017, vol. 43, no. 6, p. 325. (in Russ.)
 14. **Krinke J.** Information Flow Control and Taint Analysis with Dependence Graphs. Vancouver: Proc. of 3d International Workshop on Code Based Software Security Assessments (CoBaSSA), 2007, p. 69.
 15. **Kuck D. J., Kuhn R. H., Padua D. A., Leasure B., Wolfe M.** Dependence graphs and compiler optimizations. *Proc. of the 8th ACM SIGPLAN-SIGACT symposium on Principles of Programming Languages (POPL)*, 1981, p. 207218. DOI: 10.1145/567532.567555
 16. **Smaragdakis Y., Bravenboer M.** Using Datalog for Fast and Easy Program Analysis. In: de Moor O., Gottlob G., Furche T., Sellers A. (eds) *Datalog Reloaded. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2010, vol. 6702, p. 245251. DOI: 10.1007/978-3-642-24206-9_14

Сведения об авторах

Власенко Андрей Юрьевич, кандидат технических наук, доцент кафедры параллельных вычислений ФИТ НГУ, научный сотрудник лаборатории синтеза параллельных программ ИВМиМГ СО РАН

Мичуров Михаил Антонович, ассистент кафедры параллельных вычислений ФИТ НГУ

Царёв Василий Дмитриевич, студент ФИТ НГУ

Курбатов Максим Андреевич, студент ФИТ НГУ

Information about the Authors

Andrey Yu. Vlasenko, Candidate of Technical Sciences, Associate Proffessor of the Parallel Computations Department, Novosibirsk State University (Novosibirsk, Russian Federation)

Mihail A. Michurov, Assistant of the Parallel Computations Department, Novosibirsk State University (Novosibirsk, Russian Federation)

Vasiliy D. Tsarev, Student of the Faculty of Information Technologies, Novosibirsk State University (Novosibirsk, Russian Federation)

Maksim A. Kurbatov, Student of the Faculty of Information Technologies, Novosibirsk State University (Novosibirsk, Russian Federation)

Статья поступила в редакцию 29.12.2023;

одобрена после рецензирования 08.02.2024; принята к публикации 08.02.2024

The article was submitted 29.11.2023;

approved after reviewing 08.02.2024; accepted for publication 08.02.2024

Научная статья

УДК 519.6

DOI 10.25205/1818-7900-2024-22-1-21-30

Структура информационно-вычислительной системы для решения задачи усвоения данных при моделировании окружающей среды

**Василий Дмитриевич Котлер¹, Марина Владимировна Платонова²
Екатерина Георгиевна Климова³**

Федеральный исследовательский центр информационных
и вычислительных технологий
Новосибирск, Россия

² v.kotler@nsu.ru, <https://orcid.org/0000-0003-4359-8922>

Аннотация

В условиях быстрого роста объема информации и доступности вычислительных мощностей тема создания информационно-вычислительных систем (ИВС) становится более актуальной, чем когда-либо. В данной статье представлено описание разработанной ИВС для оценки потоков парниковых газов с поверхности Земли с помощью методов усвоения данных. Обсуждаются архитектурные и технические решения, а также специфика адаптации системы ИВС под используемую модель MOZART-4 и данные со спутника AIRS. Также освещены математические основы и представлен пример алгоритма, применяемого в конкретном случае ИВС – алгоритм ансамблевого фильтра Калмана LETKF.

Ключевые слова

информационно-вычислительная система, усвоение данных, среда моделирования

Финансирование

Работа выполнена в рамках государственного задания Минобрнауки России для Федерального исследовательского центра информационных и вычислительных технологий.

Благодарности

Авторы статьи выражают глубокую благодарность коллегам профессору Анатолию Алексеевичу Лагутину и доценту Егору Юрьевичу Мордвину за предоставленные данные и полезные дискуссии.

Для цитирования

Котлер В. Д., Платонова М. В., Климова Е. Г. Структура информационно-вычислительной системы для решения задачи усвоения данных при моделировании окружающей среды // Вестник НГУ. Серия: Информационные технологии. 2024. Т. 22, № 1. С. 21–30. DOI 10.25205/1818-7900-2024-22-1-21-30

© Котлер В. Д., Платонова М. В., Климова Е. Г., 2024

Structure of an Information and Computing System for Solving the Problem of Data Assimilation in Environmental Modeling

Vasily D. Kotler¹, Marina V. Platonova²
Ekaterina G. Klimova³

Federal Research Center for Information and Computational Technologies
Novosibirsk, Russian Federation

²v.kotler@nsu.ru, <https://orcid.org/0000-0003-4359-8922>

Abstract

The article presents the development and implementation of an information-computational system (ICS) specifically designed for assessing greenhouse gas fluxes from the Earth's surface using data assimilation methods. This system effectively interfaces with the MOZART-4 chemical transport model and utilizes satellite data from AIRS, ensuring accurate and reliable estimations of gas fluxes. At the core of this system lies the application of the ensemble Kalman filter algorithm, LETKF, for methane flux estimation, enabling the consideration of uncertainties in both data and model representations. The paper extensively discusses the architectural and technical solutions employed, emphasizing the necessity of adapting the system to the intricacies of the utilized model, thereby enhancing result accuracy. Furthermore, it elaborates on the mathematical foundations of the methodology and includes a practical demonstration of the algorithm's application within the information-computational system framework. The findings of this research hold significant value as they pave the way for further advancements in utilizing such systems for environmental modeling and data assimilation endeavors, contributing to our understanding and management of the Earth's environment.

Keywords

information computing system, data assimilation, modeling environment

Funding

The work was carried out within the framework of the state assignment of the Russian Ministry of Education and Science for the Federal Research Center for Information and Computing Technologies.

Acknowledgements

The authors of the article express their gratitude to their colleagues Professor Anatoly Alekseevich Lagutin and Associate Professor Egor Yurievich Mordvin for the data provided and useful discussions.

For citation

Kotler V. D., Platonova M. V., Klimova E. G. Structure of an information and computing system for solving the problem of data assimilation in environmental modeling. *Vestnik NSU. Series: Information Technologies*, 2024, vol. 22, no. 1, pp. 21–30 (in Russ.) DOI 10.25205/1818-7900-2024-22-1-21-30

Введение

Моделирование окружающей среды является актуальным направлением современной науки. Научные исследования в этой области позволяют не только понять процессы, происходящие в окружающей среде, но и прогнозировать их изменения в будущем. Однако, несмотря на значительные достижения в области моделирования, существуют определенные ограничения, связанные с невозможностью полного учета всех факторов, влияющих на окружающую среду. Например, математические модели не всегда могут в полной мере учитывать различные экстремальные события, такие как природные катаклизмы или антропогенные аварии, а также их последствия для окружающей среды.

Важным инструментом, используемым для оценки состояния окружающей среды, является подход, называемый усвоением данных с применением математической модели исследуемого процесса. Усвоение данных представляет собой подход, позволяющий объединить информацию из различных источников, таких как наблюдения и данные прогноза по математической модели, для получения наиболее точной оценки исследуемых параметров окружающей среды [13]. Таким образом, моделирование окружающей среды и усвоение данных играют важную

роль в научном понимании процессов, происходящих в природной среде, и могут использоваться для принятия решений в области охраны окружающей среды и устойчивого развития.

Изменению климата уделяется в настоящее время большое внимание [4]. Известно, что парниковые газы, в первую очередь углекислый газ и метан, оказывают существенное влияние на климат. Поиск и оценка источников и эмиссии парниковых газов являются приоритетной задачей для научного сообщества. Этот интерес обусловлен не только научными исследованиями, но и политическими и экономическими факторами.

Исследование эмиссии парниковых газов, в частности метана, представляет собой задачу глобального характера и имеет высокую степень сложности. Моделирование атмосферы и учет ее характеристик являются ключевыми для данного процесса. Сложность задачи неизбежно возрастает с увеличением масштаба и детализации модели атмосферы [1; 5; 6]. Исследования стимулируют создание и развитие новых моделей, а также введения в эксплуатацию большего числа новых приборов для точного измерения концентрации парниковых газов.

Внедрение современных измерительных устройств как на поверхности Земли, так и из космоса, значительно расширяет возможности сбора данных о составе атмосферы. Рост объема данных измерений наряду с созданием более точных моделей атмосферных процессов приводит к новым проблемам, связанным с хранением и обработкой информации. Для эффективного анализа этой информации и использования ее при моделировании необходимы современные информационно-вычислительные системы, способные обрабатывать, хранить и организовывать данные для проведения разнообразных экспериментов [4; 7].

Необходимость разработки информационно-вычислительной системы (ИВС) для усвоения данных обусловлена несколькими ключевыми аспектами [6]. Во-первых, с увеличением объемов данных, собираемых с различных источников, становится критически важным иметь удобные инструменты для работы с системой усвоения данных. Это включает в себя возможности эффективного доступа к данным, их обработки и оперативного анализа, а также визуализации результатов работы системы усвоения данных. Во-вторых, оптимизация и структуризация больших объемов исходных данных играют ключевую роль в успешной работе системы усвоения данных. Необходимо обеспечить эффективное хранение и управление данными, а также разработать методы и алгоритмы для их предварительной обработки и фильтрации. В-третьих, создание информационно-вычислительной системы (ИВС) является важным для научного сообщества, так как создание данного инструмента – это создание простого и понятного интерфейса для запуска моделей, проведения численных экспериментов с различными параметрами и условиями. Создание такого инструмента даст возможность ученым из смежных областей проверки своих гипотез и апробации различных климатических сценариев.

Важным аспектом в разработке информационно-вычислительных систем (ИВС) является их универсальность, т. е. способность применяться для различных математических моделей и типов данных. Также не менее важным свойством является переносимость системы на другие вычислительные средства, что обеспечивает ее функциональность и применимость в различных сценариях использования.

В связи с развитием информационно-вычислительной техники, в том числе и в нашей стране, наблюдается рост исследований в области разработки информационно-вычислительных систем для реализации задач математического моделирования, связанных с вычислениями на длительные сроки и использованием больших объемов информации. Так, в Институте вычислительной математики (ИВМ РАН) с 2005 по 2007 г. под руководством Г. И. Марчука и В. И. Агошкова выполнялся проект по разработке информационно-вычислительной системы усвоения данных, основанной на вариационном подходе, с целью анализа гидрофизических полей в океане и их расчетов [6].

В рамках исследований, проведенных И. В. Бычковым, Г. М. Ружниковым, Р. К. Фёдоровым с коллегами, были представлены два значимых проекта в области информационных технологий. Кроме того, в 2016 г. была разработана информационная система, способная пла-

нирывать и выполнять композиции веб-сервисов в гетерогенной динамической среде. Этот проект имеет большое значение для развития эффективного и гибкого использования веб-сервисов в различных приложениях, что способствует повышению производительности и удобства для пользователей [7; 8].

В статье описывается разрабатываемая авторами ИВС для задачи оценки потоков парниковых газов с поверхности Земли с применением методов усвоения данных. Реализация ИВС осуществляется для химической транспортной модели MOZART-4 и спутниковых данных AIRS. При этом на основе алгоритма ансамблевого фильтра Калмана LETKF производится оценка потоков парникового газа метана. Приводится также краткое описание задачи усвоения данных, рассматривается структура ИВС для задачи оценки потоков парниковых газов, анализируются способы реализации данной ИВС.

Задача усвоения данных при моделировании окружающей среды

Оценка состояния окружающей среды по данным наблюдений является одной из наиболее актуальных задач в настоящее время. Такая оценка проводится с привлечением прогностических моделей на основе систем усвоения данных. Под усвоением данных («data assimilation») принято понимать совместный учет данных и математической модели для наиболее точной оценки пространственно-временного распределения исследуемых величин [13]. Существует огромное количество методик усвоения данных, но, с точки зрения математической постановки задачи, все они используют один из двух подходов: вариационный или динамико-стохастический [3]. Поскольку данные наблюдений известны с ошибками, имеющими случайный характер, все методы усвоения должны учитывать статистические свойства ошибок измерений. Статистический характер имеют также так называемые «шумы» моделей, учет которых важен при проведении оценки состояния окружающей среды. Проблема учета статистических характеристик ошибок прогноза и наблюдений естественным образом решается при применении динамико-стохастического подхода (фильтр Калмана). Численная реализация фильтра Калмана для современных нелинейных моделей невозможна, поэтому в настоящее время используются различные приближения. Лидирующее положение занимает ансамблевый подход, при котором ковариации ошибок прогноза оцениваются с помощью ансамбля прогнозов по возмущенным начальным полям. Реализация ансамблевого подхода также содержит технологические сложности, связанные, в частности, с большой размерностью рассматриваемых при этом матриц [9; 10].

Изучение распространения в пространстве и времени парниковых газов, а также оценка потоков с поверхности Земли этих газов представляет одну из задач мониторинга состояния окружающей среды. Для решения этой задачи принято использовать системы усвоения данных, включающие в себя данные наблюдений и математическую модель распространения газовых составляющих в атмосфере [1; 11]. Реализация алгоритма усвоения данных для современных глобальных моделей с использованием больших объемов данных представляет собой сложную технологическую задачу, требующую разработки соответствующей ИВС.

ИВС «Оценка потоков парниковых газов по спутниковым данным»

В данном разделе статьи описана структура ИВС «Оценка потоков парниковых газов по спутниковым данным» для реализации системы усвоения данных. В ИВС используются результаты расчета прогноза по химической транспортной модели MOZART-4 в заданные моменты времени, а также спутниковые данные AIRS, в частности расчет ведется для концентрации метана.

Модель MOZART-4 (Model for Ozone and Related chemical Tracers, version 4) представляет собой трехмерную транспортную модель, разработанную для моделирования распростране-

ния атмосферных примесей, включая парниковые газы, озон и другие химические компоненты. Эта модель позволяет проводить прогнозы состояния атмосферы на различных временных и пространственных масштабах [5; 12; 6].

Спутниковые данные AIRS (Atmospheric Infrared Sounder) предоставляют обширную информацию, в том числе информацию о вертикальном профиле температуры и концентрации парниковых газов в атмосфере. Эти данные получаются с помощью специализированных приборов на борту спутников и представляют собой ценный источник информации для мониторинга и исследования атмосферных процессов [5; 12; 6].

В настоящее время в информационно-вычислительной системе используется алгоритм усвоения данных LETKF, основанный на ансамблевом фильтре Калмана (Ensemble Kalman Filter). Особенностью данного алгоритма является локальность, что упрощает вычисления. В частности, это позволяет проводить более простые вычисления и снижает вычислительную сложность системы, сохраняя при этом достаточную точность и эффективность процедуры оценки потоков парниковых газов [1; 11].

Реализация ИВС «Оценка потоков парниковых газов по спутниковым данным»

При разработке ИВС особое внимание уделяется ее архитектуре, которая должна позволять обрабатывать большие объемы данных, обеспечивать безопасное масштабирование системы и обновление функционала, а также обладать удобным пользовательским интерфейсом. Серверная часть информационно-вычислительной системы реализована на языке Python. Этот выбор обусловлен не только его надежностью и стабильностью, но и широким сообществом пользователей и разработчиков, что обеспечивает поддержку и доступ к различным ресурсам, необходимым для успешной реализации информационной системы. Python является одним из наиболее простых и распространенных языков в области научных исследований, что делает его выбор оптимальным для создания ИВС. Архитектура ИВС строится с учетом доступности модели MOZART-4 для установки и использования (исходный код модели находится в открытом доступе).

В процессе экспериментов все ансамблевые вычисления проводились с использованием стандартной linux-утилиты – xargs. Ее использование позволяет выполнять вычисления и обработку данных параллельно, что не только экономит время, но и оптимизирует использование пространства на диске. Это обеспечивает эффективность и масштабируемость работы информационной системы, что важно для обработки и анализа больших объемов данных, характерных для исследований в области моделирования окружающей среды, и усвоения данных [7]. Подробное описание информационно-вычислительной системы представлено на рис. 1.

Система состоит из нескольких ключевых компонентов, каждый из которых выполняет определенные функции в процессе обработки и анализа данных. Первым блоком является основная структура, включающая в себя элементы для передачи информации и управления выполнением внутренними процессами. Эти компоненты отвечают за передачу данных и управление процессом обработки, обеспечивая целостность и надежность системы. Далее следует блок предварительной обработки данных, где входящие спутниковые данные (спутник AIRS) и результаты вычислений прогнозов по модели MOZART-4 подвергаются предварительной обработке для дальнейшего анализа. Далее следует блок вычислений, где происходит сбор и структурирование данных и дальнейшая обработка результатов. После процедуры обработки результаты сохраняются в базе данных для общего доступа и использования. На этапе поиска оценки эмиссии парниковых газов осуществляется итеративный процесс усвоения данных с учетом различных параметров задачи усвоения данных.

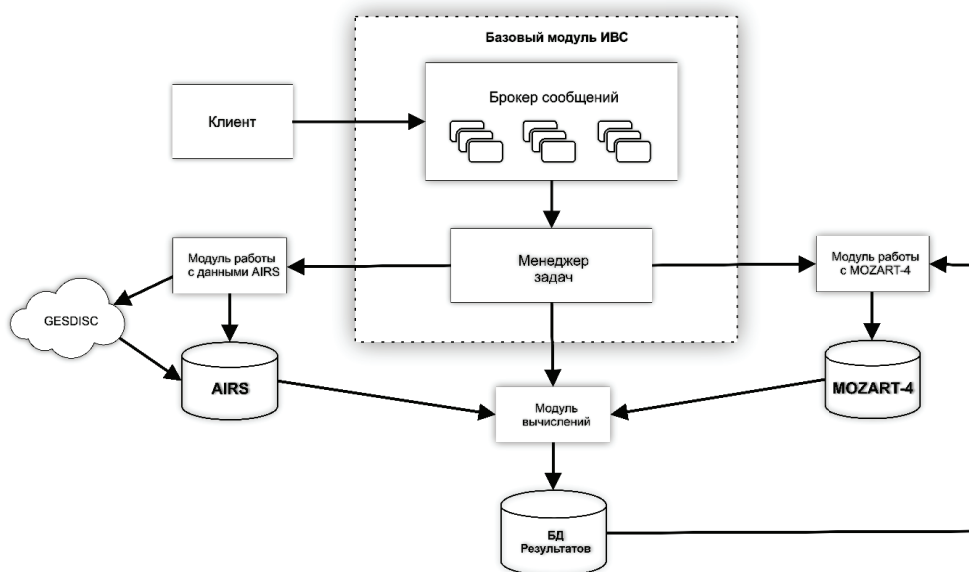


Рис. 1. Блок-схема системы «Оценка потоков парниковых газов по спутниковым данным»
 Fig. 1. Block diagram of the system “Estimation of greenhouse gas flows from satellite data”

В блоке баз данных системы усвоения данных предусмотрено хранение информации, полученной как из спутниковых наблюдений, так и результатов расчетов математической модели. Исходные данные, в том числе данные спутниковых наблюдений, и результаты расчетов поступают в ИВС в файлах с расширениями HDF (Hierarchical Data Format) и NetCDF (Network Common Data Form). В процессе работы системы на различных этапах предобработки необходимая информация считывается из этих файлов и загружается в базы данных. Это позволяет не только хранить информацию в структурированном и удобном формате, но и обеспечивает оперативный доступ к данным для дальнейшего анализа и использования. Использование баз данных обеспечивает не только возможность проведения статистических анализов за большие периоды времени, но также обеспечивает оперативный доступ к данным, что является важным аспектом в работе системы усвоения данных. Это позволяет специалистам быстро получать доступ к необходимой информации для принятия решений и анализа текущего состояния атмосферы и изменений в ней.

Важной частью процедуры обработки и анализа информации является использование математически алгоритмов усвоения данных. В разрабатываемой ИВС используется алгоритм, основанный на ансамблевом фильтре Калмана. Его использование позволяет оценивать потоки парниковых газов и предсказывать их изменения в зависимости от различных факторов, таких как изменения климата, антропогенная активность и естественные процессы [1; 10].

Следует отметить особенность реализации системы усвоения данных, связанную с использованием параллельных вычислений. Одной из ключевых характеристик используемого алгоритма усвоения данных является его локальность, что позволяет применять этот алгоритм для каждого узла сетки модели независимо друг от друга. Это обеспечивает возможность применения распараллеливания процессов и распределенных вычислений. Использование параллельных вычислений позволяет системе обрабатывать большие объемы данных более эффективно и быстро, разделяя задачи на независимые подзадачи и выполняя их параллельно на вычислительных ядрах процессора. Такой подход значительно ускоряет процесс усвоения данных и позволяет обрабатывать данные в реальном времени или в короткие сроки. Кроме того, алгоритм усвоения данных основан на ансамблевом фильтре Калмана, поэтому важной

частью реализации алгоритма является вычисление ансамбля прогнозов по модели на заданные сроки по времени по возмущенным начальным данным (порядка 20–100). Вычисление ансамблей производится независимо друг от друга и также требует параллельных вычислений для повышения эффективности алгоритма. Таким образом, возможность применения параллельных вычислений в системе усвоения данных обеспечивает большую гибкость и эффективность в обработке информации.

Центральным элементом реализации ИВС являются программные интерфейсы, через которые осуществляется взаимодействие между различными компонентами системы. Это обеспечивает эффективную передачу данных и совместную работу всех элементов. Для удобства пользователей созданы интуитивно понятные инструменты для работы с системой. Это позволяет легко управлять процессом анализа данных, вводить необходимые параметры и анализировать полученные результаты.

На рис. 2, 3 представлены страницы интерфейса для настройки процесса ИВС «Оценки потоков парниковых газов по спутниковым данным». На рис. 2 показаны параметры настройки: выбор модели, определение сетки моделирования, указание временного промежутка эксперимента и выбор типа данных для анализа и т. д. Далее эти параметры передаются в вычислительный модуль для обработки (см. рис. 3).

Рис. 2. Введение параметров

Fig. 2. Entering parameters

Рис. 3. Передача параметров в вычислительный модуль для обработки

Fig. 3. Transfer to the computing module for processing

Результаты могут быть представлены на сервере системы для дальнейшего использования или сразу визуализированы с помощью интерфейса ИВС, что обеспечивает оперативный доступ к результатам. Важно, что разрабатываемая информационно-вычислительная система обладает универсальными характеристиками. Это означает, что ее функционирование не зависит от конкретных моделей или источников данных, которые могут быть использованы в системе. В дальнейшем возможно использование в ИВС других моделей и данных измерений. Это открывает широкие перспективы для внедрения различных научных исследований и приложений, а также для адаптации системы к различным условиям и потребностям пользователей.

Благодаря клиент-серверной архитектуре разрабатываемой системы доступ к интерфейсу и клиентской части будет доступен через интернет с помощью любого браузера. После процедуры регистрации пользователи смогут создавать и добавлять в очередь процессы для обработки на сервер системы. Кроме того, серверную часть можно перенести на другие вычислительные устройства с любой операционной системой, так как она не имеет специфичных вычислительных или программных элементов. Это обеспечивает удобство использования системы, так как она может быть адаптирована под различные инфраструктуры и потребности пользователей без необходимости значительных изменений. Возможность переноса на другие ПК и легкая масштабируемость системы обеспечивают ее доступность и удобство применения для пользователей из различных точек мира.

Заключение

В статье представлена общая структура разрабатываемой ИВС «Оценка потоков парниковых газов по спутниковым данным».

На данный момент разработка информационно-вычислительной системы для анализа потоков парниковых газов с поверхности Земли прошла несколько ключевых этапов.

- Проведен тщательный анализ существующих подходов к решению основной задачи. Кроме анализа литературы, изучены методы и алгоритмы, применяемые в схожих проектах и исследованиях.
- Разработана блок-схема и архитектура информационно-вычислительной системы, включая определение основных компонентов системы, их взаимосвязей и организации процесса обработки данных.
- Разработаны алгоритмы взаимодействия между подсистемами и обмена данными между ними. Это позволило оптимизировать процессы обработки и анализа информации, а также обеспечить эффективное взаимодействие между различными компонентами системы.
- Начата работа по организации пользовательского интерфейса информационно-вычислительной системы, что включает в себя создание макетов интерфейса и разработку алгоритмов взаимодействия с пользователем, чтобы обеспечить удобство использования системы.

Все этапы разработки выполнены с учетом научных принципов и методов. Разработанные алгоритмы позволили более эффективно проводить численные эксперименты по усвоению спутниковых данных с использованием глобальной химической модели.

Список литературы

1. **Klimova E. G.** Methods of estimation of greenhouse gases concentration in the atmosphere using observations and transport and diffusion model, based on the ensemble Kalman filter // All-Russian Conference "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes, SDM 2017". 2017. Vol. 2033. P. 191–195.

2. **Bocquet M.** Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models // *Atmos. Chem. Phys. Discuss.* 2014. Vol. 14.
3. **Nakamura G., Potthast R.** Inverse Modeling. IOP Publishing Ltd., 2015.
4. **Бондур В. Г., Савин А. И.** Концепция создания систем мониторинга окружающей среды в экологических и природно-ресурсных целях // *Исследование Земли из космоса.* 1992. № 6. С. 70–78.
5. **Mordvin E. Y., Lagutin A. A.** Methane in the atmosphere of Western Siberia. Barnaul, 2016.
6. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Methane emission by petroleum industry in Western Siberia according to satellite observations Proc. SPIE 11916, 27th International Symposium on Atmospheric and Ocean Optics Atmospheric Physics, 119166N. DOI: 10.1117/12.2603438
7. **Baner P.** The digital revolution of Earth-system science // *Nature Computational Science.* 2021. Vol. 1. P. 10–113.
8. **Агошков В. И., Ботвиновский Е. А., Гусев А. В., Кочуров А. Г., Лебедев С. А., Пармузин Е. И., Шутяев В. П.** Информационно-вычислительная система вариационной ассимиляции данных измерений ИВС-Т2. М.: Изд-во СО РАН, 2008.
9. **Бычков И. В.** Инфраструктура информационных ресурсов и технологии создания информационно-аналитических систем территориального управления. Новосибирск: Изд-во СО РАН, 2016.
10. **Bychkov I., Oparin G., Tchernykh A., Feoktistov A., Bogdanova V., Dyadkin Yu., Andrukhova V., Basharina O.** Simulation Modeling in Heterogeneous Distributed Computing Environments to Support Decisions Making in Warehouse Logistics // *Procedia Engineering.* 2017. Vol. 201. P. 524–533.
11. **Evensen G.** Data assimilation. The ensemble Kalman filter, Heideberg:Springer, Verlag, Berlin, 2009.
12. **Климова Е. Г.** Экономичный алгоритм стохастического ансамблевого сглаживания // *Сибирский журнал вычислительной математики.* 2020. № 23(4). С. 381–390.
13. **Platonova M. V., Klimova E. G.** An algorithm for estimating greenhouse gas fluxes using satellite data for a global transport and diffusion model // *Journal of Physics: Conference Series.* 2021. № 1715(1). P. 012021.
14. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Regression model for reconstruction of the total methane content according to the data from AIRS hyperspectrometer and chemical transport model MOZART-4 Proc. SPIE 12780, 29th International Symposium on Atmospheric and Ocean Optics, Atmospheric Physics, 1278062.

References

1. **Klimova E. G.** Methods of estimation of greenhouse gases concentration in the atmosphere using observations and transport and diffusion model, based on the ensemble Kalman filter. *All-Russian Conference "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes, SDM 2017"*, 2017, vol. 2033, pp. 191–195.
2. **Bocquet M.** Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmos. Chem. Phys. Discuss.* 2014, vol. 14.
3. **Nakamura G., Potthast R.** Inverse Modeling. IOP Publishing Ltd., 2015.
4. **Bondur V. G., Savin A. I.** The concept of creating environmental monitoring systems for environmental and natural resource purposes. *Earth Research from Space*, 1992, № 6, pp. 70–78. (in Russ.)
5. **Mordvin E. Y., Lagutin A. A.** Methane in the atmosphere of Western Siberia. Barnaul, 2016.

6. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Methane emission by petroleum industry in Western Siberia according to satellite observations *Proc. SPIE 11916, 27th International Symposium on Atmospheric and Ocean Optics Atmospheric Physics*, 119166N. DOI: 10.1117/12.2603438
7. **Baner P.** The digital revolution of Earth-system science. *Nature Computational Science*, 2021, vol. 1, pp. 10–113.
8. **Agoshkov V. I., Botvinovsky E. A., Gusev A. V., Kochurov A. G., Lebedev S. A., Parmuzin E. I., Shutyaev V. P.** Information and computing system for variational assimilation of measurement data IVS-T2, Moscow, 2008. Publishing house of the Siberian branch of sciences, 2008. (in Russ.)
9. **Bychkov I. V.** Infrastructure of information resources and technologies for creating information and analytical systems for territorial management. Novosibirsk, Publishing house of the Siberian branch of sciences, 2016.
10. **Bychkov I., Oparin G., Tchernykh A., Feoktistov A., Bogdanova V., Dyadkin Yu., Andrukova V., Basharina O.** Simulation Modeling in Heterogeneous Distributed Computing Environments to Support Decisions Making in Warehouse Logistics. *Procedia Engineering*, 2017, vol. 201, pp. 524–533.
11. **Evensen G.** Data assimilation. The ensemble Kalman filter, Heidelberg: Springer, Verlag, Berlin, 2009.
12. **Klimova E. G.** Economical algorithm for stochastic ensemble smoothing. *Siberian Journal of Computational Mathematics*, 2020, vol. 23(4), pp. 381–390. (in Russ.)
13. **Platonova M. V., Klimova E. G.** An algorithm for estimating greenhouse gas fluxes using satellite data for a global transport and diffusion model. *Journal of Physics: Conference Series*, 2021, no. 1715(1), pp. 012021.
14. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Regression model for reconstruction of the total methane content according to the data from AIRS hyperspectrometer and chemical transport model MOZART-4 *Proc. SPIE 12780, 29th International Symposium on Atmospheric and Ocean Optics, Atmospheric Physics*, 1278062.

Сведения об авторах

Котлер Василий Дмитриевич, младший научный сотрудник

Платонова Марина Владимировна, младший научный сотрудник
Researcher ID WoS: AFI-8086-2022

Климова Екатерина Георгиевна, доктор физико-математических наук
Researcher ID WoS: S-5641-2016

Information about the Authors

Vasily D. Kotler, Junior Researcher

Marina V. Platonova, Junior Researcher
Researcher ID WoS: AFI-8086-2022

Ekaterina G. Klimova, Doctor of Physical and Mathematical Sciences
Researcher ID WoS: S-5641-2016

Статья поступила в редакцию 24.02.2024;
одобрена после рецензирования 08.04.2024; принята к публикации 08.04.2024

The article was submitted 24.02.2024;
approved after reviewing 08.04.2024; accepted for publication 08.04.2024

Научная статья

УДК 338.3:621.391:[004.5+004.67]

DOI 10.25205/1818-7900-2024-22-1-31-48

О применении кластерного анализа для развития методов оценки безопасности технических систем

Дмитрий Иосифович Лобач

Департамент по ядерной и радиационной безопасности
Министерства по чрезвычайным ситуациям Республики Беларусь
Минск, Беларусь
lobachd@yandex.ru

Аннотация

Задача исследования заключается в разработке и развитии вспомогательных аналитических подходов для осуществления экспертной деятельности. Проблема эффективной поддержки экспертной оценки безопасности актуальна на всех стадиях жизненного цикла производственного оборудования организации. Традиционные методы оценки могут сталкиваться с недостатком данных. В ходе исследований выработаны новые методы, которые, с одной стороны, должны быть удобны для цифровизации и автоматизации алгоритмов оценки, с другой – генерировать в ходе обработки данных новую информацию (проводить DATA MINING) для экспертов. В работе описаны новые возможности применения кластерного анализа, методов анализа кластеров для сферы оценки безопасности, культуры безопасности. Предложены подходы формирования кластеров данных для методов оценки безопасности производств и техники, а также сделан акцент на анализе этих кластеров с использованием теории графов. Отмечены особенности применения кластерного анализа для оценки безопасности. Даны рекомендации для развития дополнительных методов оценки безопасности с возможностью их последующей автоматизации (вычислительными средствами) в контексте цифровой трансформации экспертиз и поддержки систем управления технологических процессов.

Ключевые слова

оценка безопасности, сэйфеометрика, получение новой информации, кластерный анализ, теория графов, значимость, целевые оценки, рейтинг, культура безопасности, организационная деятельность, планирование

Для цитирования

Лобач Д. И. О применении кластерного анализа для развития методов оценки безопасности технических систем // Вестник НГУ. Серия: Информационные технологии. 2024. Т. 22, № 1. С. 31–48. DOI 10.25205/1818-7900-2024-22-1-31-48

© Лобач Д. И., 2024

On the Application of Cluster Analysis for the Development of Safety Assessment Methods for Technical Systems

Dmitry J. Lobach

Department on nuclear and radiation safety of Ministry on emergency situations
of the Republic of Belarus, Minsk, Belarus

lobachd@yandex.ru

Abstract

The objective of the study is to elaborate and develop auxiliary analytical approaches for the implementation of expert activities. The problem of effective support of expert safety assessment is relevant at all stages of the life cycle of the organization's production equipment. Traditional assessment methods may face a lack of data. In the course of research, new methods have been developed that should be both convenient for digitalization and automation of evaluation algorithms, and generate new information (conduct DATA MINING) for experts during data processing. The paper offers new possibilities for the use of cluster analysis, cluster analysis methods for the field of safety assessment, safety culture. As a result of the research, approaches to the formation of data clusters for methods of assessing the safety of production and equipment are proposed, and emphasis is placed on the analysis of these clusters using graph theory. The article highlights the features of the use of cluster analysis for security assessment. Recommendations are given for the development of additional methods of safety assessment with the possibility of their subsequent automation (by computers) in the context of digital transformation of expertise and support of control systems of technological processes.

Keywords

safety assessment, safeometrics, data mining, cluster analysis, graph theory, significance, target assessments, rating, safety culture, organizational activity, planning

For citation

Lobach D. J. On the application of cluster analysis for the development of safety assessment methods for technical systems. *Vestnik NSU. Series: Information Technologies*, 2024, vol. 22, no. 1, pp. 31–48 (in Russ.) DOI 10.25205/1818-7900-2024-22-1-31-48

Введение

Проектирование, строительство, изготовление, эксплуатация производственных объектов и установок сопровождаются проведением ряда плановых и внеочередных оценок безопасности в интересах разных заинтересованных сторон. Результат проведения таких экспертных задач зависит от квалификации экспертов, методов сбора и обработки данных. Для формулирования и решения задач с дифференцированным подходом с количественной оценкой уровня или индекса безопасности и формирования вычислительного аппарата сэйфеометрики можно использовать методы обработки количественных закономерностей и числовых массивов в сочетании с кластерным анализом [1; 2].

Существующие подходы оценки безопасности – детерминистические и вероятностные – различаются по своим целям и группам анализируемых параметров. Их использование основано на обработке экспериментальных и эксплуатационных данных практического применения объектов и установок.

Современный анализ протекания комплексных технологических процессов невозможно представить без применения методов математической обработки и визуального отображения информации. Зачастую использование традиционных подходов не дает прорывных результатов при анализе и требуется развивать методологическую основу экспертной деятельности, во многих случаях установившегося в настоящее время порядка математической обработки данных не имеется. Результаты наблюдений и систематической надзорной работы представляют собой представительные массивы численной информации. Надлежащая их обработка должна привести к должному теоретическому осмыслению и установлению закономерностей,

позволяющих прогнозировать аспекты безопасности. В кластерном анализе для обработки информации получило распространение применение закономерностей из теории графов.

В работах [1; 3] сформулированы пояснения о развитии дополнительных методов оценки уровня безопасности, в том числе с использованием концепции сэйфеометрики.

Для построения теории и описания области технического познания и оценки вычисления непосредственно ненаблюдаемых величин на основании обработки априорной информации и эмпирических данных, используя подходы сэйфеометрики, необходимо ориентироваться на подходящие математические методы и функциональные зависимости, как об этом говорилось в [2]. Основное свойство модели должно заключаться в ее способности прогнозировать уровень безопасности или его изменение.

Применение дополнительных методов хотя и увеличивает затрачиваемое на экспертные оценки время, но дает дополнительные сведения для оценки данных в ходе проведения DATA MINING (получения новой информации в ходе обработки данных). В связи с этим важно также иметь возможность цифровизации аналитического аппарата подходящих и удобных для использования методов. Это также повышает наглядность восприятия результата, что и предлагается в настоящей работе в разрезе применения теории графов для развития методов оценки безопасности. При наличии у специалистов клипового мышления [4] дополнительные методы могут задублировать рассмотрение, что позволит получить более достоверный результат.

Методология анализа данных

Исследователь (эксперт) должен определить цель и задачи анализа и оценки. Исходя из этого он формирует совокупность входных данных, осуществляет начальный рейтинг параметров. Ход анализа будет определяться его целью. Полнота и круг рассматриваемых вопросов анализа определяются его задачами.

В ходе деятельности по регулированию вопросов безопасности необходимо определить приоритет и частоту надзорных мероприятий, сделать акцент на внимание к отдельным системам и оборудованию из их совокупности, определить критическое состояние оборудования или систем, а также иметь возможность определять предвестники инцидентов и аварий.

При проведении исследования необходимо учитывать существенные факторы, влияющие на состояние безопасности, и отсеять (отделить) факторы, которые не оказывают влияние на состояние или не имеющие значения на определение причинно-следственной связи для оценки уровня безопасности. Исключение ничтожных факторов производят на первичной основе. Указанное отсеивание обычно проводится на основе ранжирования в порядке убывания или возрастания установленного признака. Ранжирование факторов представляет собой их расположение в последовательности, например, в порядке убывания их влияния на значимые события в системе. Для ранжирования можно использовать подходящие математические закономерности, например, теорию графов [5].

Нормативные и эксплуатационные документы позволяют составить и получить перечни установленных и эксплуатационных показателей – индикаторов безопасности (физические величины, вероятности, риски), а также иных характеристик производственного процесса. Эти индикаторы составляют массивы данных для анализа и оценки безопасности $A_i = 1...z$, каждый из которых группируется из отдельных тематических элементов. Компоненты разных массивов будут иметь связь между собой в зависимости от цели анализа. Они являются исходными параметрами для дифференцированного анализа безопасности оборудования и устройств, в отношении которых проводится рассмотрение. Каждый массив A_i представляет собой последовательность однородных характеристик с акцентом принадлежности к группам систем и оборудования, классам безопасности, группам систем оборудования, последствиям изменений уровня безопасности, содержанию и серьезности нарушений и т. д. В рассматриваемых

массивах нумерация параметров должна соответствовать их рейтингу (важности) по влиянию на безопасность или тяжесть последствий: в начале массивов располагаются наиболее значимые параметры, в конце – менее значимые. Соседние элементы массивов могут быть связаны между собой только по рейтингу. Однако такие элементы могут иметь между собой связь через один или несколько элементов иных массивов (одного или нескольких), формировать совокупность их причин или признаков, отдельную логическую последовательность данных, формировать первичные кластеры для анализа.

В экспертной деятельности применяются чередование обработки и анализа полученных сведений, конечной целью которых является оценка изменения уровня безопасности как функции $q(X)$ от одного (X) или нескольких ($X_1, \dots, X_Y$) проектных или эксплуатационных параметров. Такие функции могут быть стационарными или нестационарными. Применение разнообразия в методах оценки безопасности придает результату весомость и значимость для последующего принятия управляющих решений.

Полученная функциональная зависимость величин фактически является математической моделью, которую удобно использовать с целью получения сведений об объекте анализа и способной реализовываться для решения программно-алгоритмическими средствами. К моделям предъявляются требования в способности прогнозировать значение результата с установленной точностью. Однако для косвенно определяемых величин точность может быть оценена только в ходе практической проверки на основании фактически наблюдаемых величин. По результатам такой проверки можно судить о том, является ли модель адекватной, т. е. удовлетворяющей первоначальной гипотезе эксперта.

Первым этапом анализа данных является применение кластерного анализа с целью выявления приоритетных областей внимания, исходя из целей анализа, например, анализ изменения параметров для групп определенного оборудования, поиск факторов риска для аварийных ситуаций, формирование недостаточно обследованных областей производств и т. п. Применение кластерного анализа является поисковым подходом системы DATA MINING, который позволяет структурировать массивы эмпирических данных, сформировать однородные, удобные для исследования группы или кластеры. Фактически в ходе применения кластерного анализа совместно с ранжированием решается задача классификации, стратификации данных и выявления однородных мажоритарных и доминантных структур параметров.

Для проведения кластеризации фактов при современном уровне развития науки и техники можно использовать как автоматические (программируемые модели), так и неавтоматические (визуальные, непрограммируемые модели) подходы [6].

Степень влияния как воздействия установленного кластера, так и группы факторов на уровень безопасности может изменяться во времени. Однако если это не имеет явной технической причины, например, из-за метеорологического влияния на оборудование и системы, то можно предположить, что причину следует искать в формировании совокупности индивидуальных, групповых или организационных элементов человеческого фактора. Применение кластерного анализа позволяет использовать интеллектуальную группировку факторов для определения оптимального или менее затратного способа планирования работ, определив заведомо релевантное подмножество факторов риска.

Кластерный анализ при оценке безопасности может обнаруживать устойчивые или спорадические связи влияний. Однако во всех случаях это будут достоверные эмпирические результаты. В разрезе поиска «слабых» звеньев проекта следует уделять внимание случаям устойчивого формирования кластеров влияния на уровень безопасности, устойчивую типологию. Такие закономерности необходимо учитывать для приоритетного внимания при проверках состояния безопасности или проведении экспертных рассматриваний, определения численности вовлеченных в эту работу сотрудников.

На втором этапе анализа требуется проранжировать сами сформированные кластеры. Ранжирование кластеров даст возможность определить окончательную приоритетность к группам внимания.

В настоящее время уже существуют разные методы, которые имеют возможность цифровой обработки и представления данных – индикаторы безопасности атомных станций [7], 9 индикаторов в Канаде и 7 направлений в США [8], 5 классов опасности в России [9].

В ходе проведения кластерного анализа проводится обработка многомерных массивов статистических данных различной размерности [10–13]. В результате выделяются группы объектов, однородных по рассматриваемым признакам, непересекающиеся подмножества, проводится выявление «сгущения» объектов внимания. К каждой отдельной группе признаков (кластеру) можно применять дифференцированные управленческие решения. При применении ранжирования к факторам процедуры их различная размерность и многомерность массивов не являются помехами для проведения экспертного анализа. Аналитик сам может формировать интересующие группы факторов и кластеры. Несмотря на имеющиеся размерности у величин, кластерный анализ позволяет делать этот процесс наглядным. На основании типологии влияющих на безопасность факторов аналитик может исследовать полезные для своей работы схемы группирования объектов, наглядно представить и проверить свои гипотезы. Достоинствами кластерного анализа являются типология данных, сжатие объемов информации для анализа, DATA MINING для обнаружения новых данных, скрытой информации, обнаружения структуры в изучаемой совокупности объектов, построение новых классификаций для слабоизученных явлений.

В ходе кластерного анализа обычно выделяют следующие основные проблемы: различная размерность используемых для анализа показателей и многомерность исследуемых объектов.

Наиболее распространенным методом исследования кластера является вычисление на основании изменений величин однородных факторов выборочной дисперсии или выборочного корреляционного момента, либо входящих в состав кластера так называемых метрик или расстояний, например, евклидова метрика, хеммингово расстояние. Однако для ранжирования самих кластеров необходимо использовать характеристики не факторов, а именно кластеров, которые могут быть рассмотрены в виде графов.

Для решения и формирования задач с дифференцированным подходом и количественной оценкой уровня или индекса безопасности и формирования вычислительного аппарата сэйфеметрики можно использовать количественные закономерности обработки массивов данных кластеров из теории графов [1].

Применение анализа кластеров с использованием теории графов для оценки безопасности технических систем

Для применения теории графов исследователь должен построить граф G с вершинами на основании интересующих его структурных элементов массивов $A_i = 1...z$, элементы внутри которых должны быть выстроены в порядке убывания рейтинга (значимости), от более высокого уровня к менее значимому элементу. Исходя из сформулированных задач, графов может быть несколько.

Суть методов заключается в определении изменений одного или нескольких параметров графа (количество вершин (V) и связей (ребер) (E), валентность (степень) вершин ($deg(v)$) и т. д.). С учетом изменения параметров графа анализируется наличие экстремумов, определяются критические системы, их сложность и состояние, делается вывод о приоритетных эксплуатационных или регулирующих действиях. Для количественного анализа сравниваются как значения показателей безопасности, так и параметры графов.

Порядок графа $|G|$, т. е. количество его вершин, будет определять сложность и комплексность рассматриваемой системы. Чем сложнее система, тем большее время необходимо затратить на ее осмотр, изучение ее свойств, например, безопасности. Предположим, что периодичность (период) этих действий $T \frac{1}{|G|}$. Количество связей в графе $\|G\|$ отражает картину сложности системного взаимодействия, для контроля которого требуется больше внимания, действий, частый осмотр и поэтому также периодичность $T \sim \frac{1}{\|G\|}$.

Очевидно, что более значимые системы требуют большего внимания (меньшая периодичность осмотра, более частые действия). Их важность или значимость $S \frac{1}{T}$. Поскольку сумма степеней вершин (валентность вершин) графа равна удвоенному числу его ребер $\sum_{v \in V} \deg(v) = 2|E|$, то $S \sim 2|E|$.

Аналитическое описание методов для проведения целевых оценок

Цель применения кластеризации при анализе безопасности – провести группирование внимания, определить для анализа сосредоточение признаков (характеристик, изменений контролируемых величин и др.) в кластере или кластерах либо отсутствие такого сосредоточения.

Построение кластера. Эксперт определяет исходный массив данных для построения кластера, ту тему, которой он хочет уделить свое основное внимание. Массивы данных описывают разные аспекты структуры и свойств технических систем (ТС) и управления, например, перечни ТС, эксплуатационные данные, характеристики системы управления ТС, вовлеченного в управление персонала. Элементами массивов могут быть параметры ТС, их изменение, количественные и качественные характеристики управления, его персонала, количество вовлеченных в деятельность с ТС специалистов. Для упрощенного построения кластера важно само наличие характеристик ТС, а для более детального – изменение параметров и наличие связей между параметрами и изменениями. Последовательно из каждого массива A_{i+1} выбирается один элемент, который изменился или вообще используется для непосредственной характеристики элемента предыдущего массива A_i или элемента основного критерия оценки из A_0 (если он принят в подходе или взят экспертом для анализа). В неавтоматизированном режиме эксперт сам определяет степень важности установления связи (ранга или рейтинг). Для всех элементов одинаковой важности тоже могут быть установлены связи в кластере при фактическом наличии таковых.

Визуально построение кластера из элементов A_i для его последующего анализа аналогично наглядности структуры построения связей в корреляционных метриках или расстояниях. Однако различия заключаются в конкретном замысле эксперта и смысле результата его анализа для целей оценки безопасности ТС. Схемы построения кластеров в целях применения методов для целевых оценок приведены далее в статье на основе примеров в разрезе описания алгоритма оценки безопасности ТС для системы поддержки принятия решений (СППР).

Рейтинг первого уровня (рейтинг признаков). Цель – определить приоритет среди разных аналитических задач или экспертных точек зрения на состояние оборудования и систем по отсутствию достаточного количества данных или доступному количеству изменения признаков или параметров.

Для каждого кластера экспертной проблематики конструируются графы $G_{i=1...N}$ и каждому из них определяется условное значение значимости S_{G_j} . Набору графов будет соответствовать набор из величин значимостей, в котором можно определить максимальное и минимальное

значение, которое укажет на область недостаточного тактического внимания в управлении безопасностью.

Ряд из значимостей выстраивается в порядке их убывания (или возрастания, в зависимости от замысла эксперта). Определяются графы с наибольшей значимостью. Графы с наименьшей значимостью анализируются на достаточность количества данных. При минимальной значимости и недостаточном количестве данных следует считать область анализа малоизученной и требующей дополнительного внимания (так называемая GAP-область анализа).

Отличие применения анализа рейтинга первого уровня от применения корреляционной метрики или расстояния состоит в другом аналитическом подходе. В рейтинге признаков делается попытка оценки непосредственно важности или значимости, а также требуемой периодичности осмотра. При анализе корреляционных метрик или расстояний (при изучении изменений характеристик) существует проблема понимания смысла и размерности вычисленной величины, которая для удобства анализа рассматривается как относительная (в относительных единицах).

Рейтинг второго уровня (рейтинг задач). Цель – определить, какие задачи анализа безопасности являются наиболее приоритетными при рассмотрении.

Проводится оценка адекватности постановки задач нескольких типов, для которых была проведена процедура рейтинга первого уровня. Формулируется несколько видов задач на основании элементов массивов $A_i = 1...z$. Для каждого A_i однотипной задачи из рассматриваемых массивов конструируется одинаковое количество графов $G_j = 1...N$, на основе элементов A_{ij} , для каждого графа определяется значимость S_{G_j} .

Рейтинг задачи определяется суммированием значимостей графов задачи $S_i'' = \sum_j S_{G_j}$.

Строится упорядоченный ряд значимостей задач в порядке убывания. Определяются наиболее важные задачи.

Чем больше элементов с высоким рейтингом, тем больше значимость задачи, необходимо анализировать такое «сгущение» признаков или событий.

Рейтинг третьего уровня (рейтинг внимания). Цель – оценить организационную работу и планирование деятельности в зависимости от того, где уделяется больше или меньше внимания – конкурирующим по вопросам безопасности ТС или оборудованию, а также, где имеется или установлено больше или меньше признаков для контроля оборудования (стратификация признаков).

Выполняется подбор нескольких видов задач, как для рейтинга задач, но количество графов для каждой задачи может быть произвольным (k_i). В целом, где меньше признаков в массиве, там требуется бóльшая концентрация внимания, так как имеется менее выделенная их стратификация.

Проводится рассмотрение удельной величины для концентрирования стратегического внимания на задачах в области безопасности, а также в соответствии с признаками массивов

$A_i = 1...z$ после проведения анализа задач $S_i''' = \frac{S_i''}{k_i}$, где k_i – количество рассматриваемых случаев задач одного типа $i(k_{i_{MAX}} = j_{MAX} = N)$. Определяются величины S_i''' , которые выстраиваются в порядке убывания.

Максимальное значение значимости показывает, что там уделяется бóльшее внимание или только там проводится рассмотрение из перечня признаков для выбора. Предположительно недостатки организационной работы будут незначительными, если нет явно выраженных максимальных или минимальных величин S_i''' (когда различия не превышают значений $\pm 10\%$ или иное значение, по мнению эксперта). В случаях минимального значения S_i''' очевидно имеется существенный недостаток внимания в работе.

Возможный случай построения экспертом своей работы. Методы могут использоваться как по отдельности, так и совместно, в комбинации. Каждому перебору параметров для анализа строится граф G с величиной значимости S . При построении последовательности величин значимостей по максимальному и минимальному значению можно проанализировать сильные и слабые стороны системы организации деятельности по обеспечению безопасности. Наличие нескольких графов с экстремальными величинами значимости может указать на недостатки в системном планировании работ.

Экстремумы для графов определяются с целью понимания анализируемых данных как максимумов (зон повышенного внимания или реальной ситуации), так и минимумов (зон отсутствия внимания, «разрывов» в системной деятельности) для дальнейшего их анализа. Важно учесть возможность анализа ресурсов организации для эксплуатационных или контрольных целей. Далее экспертом производится поиск новых кластеров и типизация их признаков.

Пусть взяты выборки из массивов данных для анализа и оценки безопасности, где различаются первый приоритет i_1 и второй приоритет i_2 . Для первого приоритета максимальное количество случаев задач одного типа $k_{i_1 \text{ MAX}}$, для второго – $k_{i_2 \text{ MAX}}$. Если $k_{i_2 \text{ MAX}} > k_{i_1 \text{ MAX}}$, то детализация рассмотрения для второго приоритета больше, чем для первого. Когда $k_{i_2 \text{ MAX}} = k_{i_1 \text{ MAX}}$, а $S_{i_1}^{\text{III}} = S_{i_2}^{\text{III}}$, то на практике имеет место равномерное распределение внимания в работе, все аспекты для анализа и оценки безопасности охвачены.

Если для анализа взято $j = k_{i \text{ AN}}$ из диапазона $1 < j < N$, то можно рассмотреть комбинации случаев из табл. 1.

Таблица 1

Комбинации случаев

Table 1

Combinations of cases

Приоритет <i>Priority</i>	Критерий второго уровня <i>Second-level criteria</i>	Критерий третьего уровня <i>Third-level criteria</i>	
первый приоритет, i_1	$S_{i_1}^{\text{II}} > S_{i_2}^{\text{II}}$	$S_{i_1}^{\text{III}} > S_{i_2}^{\text{III}}$	$S_{i_1}^{\text{III}} < S_{i_2}^{\text{III}}$
второй приоритет, i_2	$S_{i_1}^{\text{II}} < S_{i_2}^{\text{II}}$	не все рассмотрены аспекты задач для i_2	не все рассмотрены аспекты задач для i_1

В ходе анализа количество рассматриваемых случаев может образовывать зоны (области) M (случаи частично учтены, фактически рассмотрены) и N (все учтено, рассмотрено максимально возможное) в диапазоне $1 < M < k_{i \text{ AN}} < N$, где $M < k_{i \text{ F}}$ является количеством случаев, которые могут быть рассмотрены при фактическом профессионализме работников. Если проявляемый ими профессионализм максимален (100 %), то $k_{i \text{ AN}} = k_{i \text{ MAX}}$. Если $k_{i \text{ AN}} = k_{i \text{ F}}$, то эксперт сужает область рассмотрения, и это должно быть либо обосновано, либо свидетельствует о недобросовестности эксперта.

Для оценок будет иметь место соотношение $S_{i \text{ F}}^{\text{II}} = \sum_{j=1}^M S_{G_j} < \sum_{j=1}^N S_{G_j} = S_{i \text{ MAX}}^{\text{II}}$. Величина $S_{i \text{ MAX}}^{\text{II}}$ зависит от уровня профессионализма, с его повышением увеличивается уровень внимания и глубины рассмотрения задач.

Уровень культуры безопасности и фактического профессионализма будет характеризоваться соотношением рассмотренных величин $\frac{S_{i \text{ F}}^{\text{II}}}{S_{i \text{ MAX}}^{\text{II}}} = R_{i \text{ SC}}^{\text{II}}$, т. е. компонентов для задачи

одного типа i . При $M \rightarrow N$ значение $R_{iSC}^{II} \rightarrow 1$, т. е. уровень культуры безопасности (от англ. *Safety Culture*) повышается до максимального (рис. 1).

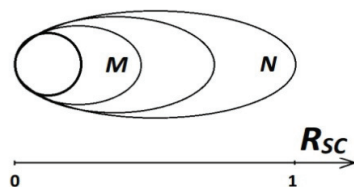


Рис. 1. Схема оценки уровня культуры безопасности по относительной шкале на основании величины зоны (области) данных для задач

Fig. 1. Scheme for assessing the level of safety culture on a relative scale based on the size of the zone (area) of these tasks

В ходе анализа можно определить реперные аналитические точки для рассмотрения организационной деятельности и планирования, выработки рекомендаций на основании табл. 2.

Таблица 2

Примеры реперных аналитических точек

Table 2

Examples of reference analytical points

Реперная точка <i>Reference point</i>	Значение <i>Meaning</i>
$\frac{S_{iMAX}^{II}}{k_{iMAX}} = S_{iMAX}^{III}$	потенциальное значение; ситуация, когда использованы и вовлечены максимальные средства и внимание
$\frac{S_{iF}^{II}}{k_{iF}} = S_{iF}^{III}$	фактическое значение; если задачи непредставительные для состояния оборудования или безопасности, то $S_{iF}^{III} < S_{iMAX}^{III}$; если $S_{iF}^{III} > S_{iMAX}^{III}$, то уменьшение внимания $k_{iF} < k_{iMAX}$ не сказывается на уровне состояния оборудования или безопасности
$\frac{S_{iF}^{II}}{k_{iMAX}}$	доля рассмотренных данных, вовлеченных в рассмотрение задачами k_{iF} из k_{iMAX} возможных, соответствующая уровню R_{iSC}^{II} отношения к работе задач i
$\frac{(S_{iMAX}^{II} - S_{iF}^{II})}{k_{iMAX}}$	доля упущенных данных; полученное значение показывает величину зоны (области) W , т. е. $W = N/M$, области недостатков организационной деятельности и планирования

Тем не менее если для работы или анализа не используется вся доступная детализация массива параметров, то недостаточное внимание должно быть или обосновано, или скомпенсировано другими подходами рассмотрения. Если присутствует несовпадение зон (областей) M и N , то может иметь место сознательное снижение уровня качества деятельности.

Метод произвольной задачи. Цель – проанализировать сильные и слабые стороны системы организации деятельности по обеспечению безопасности по выбранным основным критериям признаков.

Из массивов признаков $A_{i=1...z}$ выбирается массив основных критериев оценки A_0 , а в нем исходный элемент для анализа. Проводится поиск резервов путем непрерывного перебора связей между элементами массивов при поступлении или изменении для них данных. В кластере элементы последующих массивов должны соответствовать предшествующим, иметь с ними причинно-следственную связь. Порядок рассмотрения элементов массивов – ранжирование (согласно замыслу эксперта) для анализа массивов или критериев анализа по значимости. Основные критерии оценки меняются от 1 до Z для каждой оценки, и далее анализ непрерывно (например, мэйнфреймом или суперкомпьютером) повторяется сначала, как показано в иерархическом виде схематично в табл. 3 (критерии анализа – вершины графа, критерии в соседних ячейках могут образовывать связи в графе; построения, рассмотрения и анализ кластера производится в направлении стрелок).

Таблица 3

Примерная схема построения, рассмотрения и анализа кластера

Table 3

An approximate scheme for building, reviewing and analyzing a cluster

Действия эксперта <i>Expert actions</i>	Ранжирование (замысел эксперта) для анализа массивов или критериев анализа по значимости (глубина, фактическое рассмотрение увеличивается с рангом) →					
порядок (этапы) рассмотрения массивов →	1→ (основной критерий оценки)	2→	3→ ↓	4→	5→	...
построение гра- фа кластера →	A_0	A_1	A_3	A_6	A_5	...
			A_4			...
		A_2	A_7	A_8		...

Для каждого из построенных графов $G_{j=1...z}$ определяется значимость S_{G_j} , из величин которых получается последовательность от больших значений к меньшим. Графы с большей величиной значимости относят к сильной стороне работы, а с меньше – к слабой.

Метод азимутальных годовых (временных) диаграмм. Азимутальные диаграммы можно использовать для формирования временных кластеров событий (связанных по времени или промежутку времени). Значение величины удобно откладывать в полярной системе координат. Для оценки условно 365 или 366 дней следует сопоставить 360 градусам (5 или 6 дней исключаются систематическим путем, например, из месяцев с 31 днем). Можно рассмотреть для анализа любой другой временной промежуток до 360 дней. Путем наложения и (или) поворота диаграмм разных параметров, сравнивая величины в разные промежутки времени (годы), можно определить приоритетные отрезки времени, чтобы уделить в этот временной промежуток внимание вопросам с более высоким рейтингом (рис. 2).

По отметке на вертикальной линии можно произвести настройку времени анализа (сдвиг или совмещение) рассматриваемых функций или признаков ТС (совмещение дат исходных событий, дат экстремумов, иных технологически важных календарных дат и т. п., а также характеристик кластеров при рассмотрении диаграмм для них). Таким образом может быть рассмотрен кластер, сформированный синхронизацией или совпадением времени некоторых событий. Применение диаграмм позволяет получить дополнительные элементы для анализа кластеров или графов, которые сформировал эксперт согласно своему замыслу (по углу φ по-

лучить время события t или по изменению углов $\Delta\varphi = \varphi_2 - \varphi_1$ – промежуток времени Δt между событиями).

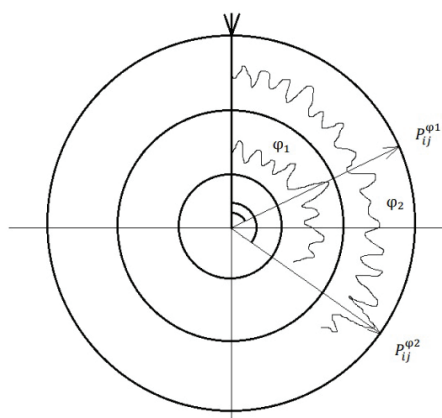


Рис. 2. Схема применения азимутальных годовых (временных) диаграмм для получения связи событий (по времени или промежутку времени)

Fig. 2. Scheme of application of azimuthal annual (time) charts to obtain the relationship of events (by time or time interval)

Для рассмотрения событий можно выбрать промежутки времени (интервал поиска, промежуток анализа или внимания) от начала отсчета до пика величины события (период накопления влияющих факторов) или промежуток между пиками событий, изучая факторы консервации последующего события.

О расстановке акцентов при анализе кластеров для оценки безопасности ТС, специфика применения существующих методов

Деятельность эксперта не хаотична. Он проводит целенаправленную работу по поиску важных элементов для оценки безопасности, чтобы сделать свой вывод или дать рекомендации.

Для анализа кластеров можно использовать имеющийся методологический арсенал из сферы экономики и исследования рынков, но в новом контексте. В разрезе оценки безопасности надо сделать соотнесение категорий и понятий, их ролей в целях изучения уровня безопасности. Математический поиск параметров графов, определение минимумов и максимумов, рейтингов и проведение ранжирования требуют тщательного анализа структуры факторов и картины данных, необходимо делать выводы и проводить DATA MINING. После проведения адекватного понятийного транслирования можно дополнительно использовать некоторые, уже хорошо разработанные методы анализа.

Метод GAP-анализа (от англ. *Gap* – разрыв) представляет целенаправленный поиск «разрывов» в понимании «слабых» звеньев и сторон структур и систем в разрезе обеспечения безопасности и при принятии решений по вопросам управления [2]. Математически это осуществляется в ходе поиска экстремумов данных. GAP-анализ является методом, с помощью которого осуществляется поиск шагов для достижения цели повышения безопасности путем устранения «слабых» звеньев, элементов. В ходе этого анализа надо делать выводы о текущем состоянии ситуации, идентифицировать «разрывы» между текущим и желаемым или требуемым результатом. В случае проведения анализа причин такого «разрыва» (между текущим и желаемым или требуемым результатом) можно разработать план компенсации мероприятий (по достижению желаемого результата) и выработать рекомендации, например, для уменьшения влияния человеческого фактора, несанкционированных действий и т. п.

Главная цель GAP-анализа стратегических утечек, «разрывов» тут заключается в том, чтобы выявить те возможности, которые могут стать для организации эффективными преимуществами в вопросах обеспечения безопасности, выделить стратегические недостатки в обеспечении безопасности или контроле за ее уровнем.

Метод SWOT-анализа (от англ. *Strength* – сила, *Weakness* – слабость, *Opportunity* – возможность, *Threat* – угроза) в наших целях тут может быть использован для расстановки приоритетов, поиска резервов и слабых сторон организации в разрезе анализа вопросов обеспечения безопасности как с описательными задачами, так и для получения рекомендаций. В целом, в ходе применения этого метода осуществляется структурирование, типизация и оценивание информации, что позволяет выделить сильные и слабые стороны организации деятельности, а также возможные сопутствующие негативные или позитивные факторы, либо «эффекты» организации эксплуатационной или контрольной деятельности в широком спектре вопросов.

Аспекты SNW-анализа (от англ. *Strength* – сила, *Neutral* – нейтральный, *Weakness* – слабость) эксперт может применять для анализа системных вопросов безопасности при организации конкретной работы в организации. Анализ сторон как эксплуатирующей, так и контрольной организации на основании результатов экспертной и надзорной деятельности проводится в ходе анализа организационной культуры в управлении и персональной культуры безопасности для поиска резервов системы (запаса безопасности).

SNW-анализ – это анализ сильных, нейтральных и слабых сторон организации, который позволяет определить, что удалось достичь и чего можно достичь. Правдивая, объективная оценка и анализ дадут при изучении перспективные выводы и рекомендации. Признаки планирования фактически отражают существующий ресурс и возможности структуры организации. Как показала практика, для стратегического анализа внутренней среды организации в качестве нейтральной позиции можно фиксировать состояние для текущей ситуации и относительно ее проводить все оценки.

SNW-анализ часто нужен экспертам по вопросам безопасности для разработки стратегии планирования или деятельности, так как позволяет более глубоко оценить ситуацию и провести необходимые мероприятия и исследования внутренних ресурсов, без которых невозможно полноценно и продуктивно действовать. В этом случае эксперту надо определить факторы оценки, определить средние показатели для установления ориентиров или масштаба оценки (определить критерии положительных и отрицательных признаков), оценить организацию по каждому фактору и выработать рекомендации.

Алгоритм оценки безопасности ТС для СППР

На основании описания перечисленных методов для проведения целевых оценок можно сформулировать исходные элементы внедрения поэтапного подхода оценки безопасности ТС. Количественные характеристики свойств ТС, результаты их анализа могут быть использованы как дополнительные входные данные в СППР.

Особенности рассмотренных в статье методов анализа заключаются в возможности их применять:

- для формирования поэтапного алгоритма оценки на основе аналитического аппарата кластерного анализа;
- внедрения в цифровую поддержку экспертиз и автоматизированных систем управления технологическим процессом (АСУ ТП), СППР для повышения качества экспертиз и автоматизированного анализа.

Для АСУ ТП важно иметь в наличии всестороннюю текущую информацию и данные о состоянии ТС для проведения объективного анализа, контроля и принятия управленческих решений. Предлагаемые в статье методы позволяют получить для этого вспомогательную визуаль-

ную информацию. Представленные подходы анализа, такие как разновидность DATA MINING, являются элементами информационного поиска и ситуационного анализа, которые могут применяться в ходе непрерывного или дискретного поступления данных в АСУ ТП или СППР.

Для проведения плановых работ требуется определить количество направлений деятельности и полноты (объема) анализа в цепочке работ «рейтинг 1» → «рейтинг 2» → «рейтинг 3» для конечного числа имеющихся ТС, для количества проверяемых ТС или работающих с ними специалистов, а также в отношении количества вовлеченных для экспертизы сотрудников (по нарастанию их привлечения). Это необходимо, чтобы избежать переполнения информацией при работе и для корректного анализа. Для целевых работ один и тот же метод целесообразно применять с одинаковой полнотой к разным направлениям оценки безопасности ТС.

Таблица 4

Таблица приемлемости применения методов настоящей статьи
для практических целей анализа безопасности ТС

Table 4

The table of acceptability for the methods of this article
for practical purposes for safety analysis of TS

Виды работ, исполнители работ, виды ТС <i>Types of work, performers of work, types of TS</i>		Плановая работа <i>Planned work</i>			Целевая работа <i>Targeted work</i>	
		Рейтинг первого уровня (рейтинг признаков)	Рейтинг второго уровня (рейтинг задач)	Рейтинг третьего уровня (рейтинг внимания)	Метод азимутальных годовых (временных) диаграмм	Метод произвольной задачи
Инспекции, проверки	плановые	да	—	—	Синхронизация важности мероприятий по времени их проведения для разных объектов, специалистов, служб	Таргетированное внимание при выборе проверки объектов
	целевые	да	возможно	—		
	комплексные	да	возможно	возможно		
Работники, службы	специалисты	да	—	—		
	руководители	да	возможно	—		
	группы управления или планирования	да	возможно	возможно		
ТС	отдельное оборудование	да	—	—		
	отдельная система	да	возможно	—		
	цех, участок с комплексом ТС	да	возможно	возможно		

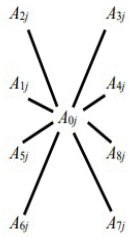
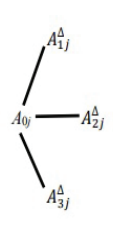
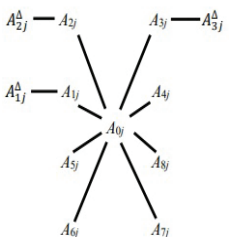
Применение алгоритма оценки безопасности ТС в целях цифровизации работы эксперта или в контексте функционирования СППР возможно при автоматизации вычислительными средствами процедур перебора подходящих данных (элементов) массивов для их анализа и построения функциональных кластеров. Экспертная система может быть представлена в виде интерактивной оболочки для консультантов или лиц, принимающих решения, чтобы помочь им собирать полезную информацию из комбинации данных, проводить DATA MINING

Таблица 5

Алгоритм-схема построения кластеров в целях их использования
для оценки безопасности ТС (с иллюстративными случаями)

Table 5

The algorithm-scheme for building clusters in order to use them
to safety assessment of TS (with illustrative cases)

Структура входных данных, массивы, ТС, индикаторы безопасности <i>Input data structure, arrays, TS, safety indicators</i>				Кластер систем и комплексов <i>Cluster of systems and complexes</i>	Кластер изменений <i>Cluster of changes</i>	Смешанный кластер <i>Mixed cluster</i>
Признаки массивов			Элементы массивов	Подзадача 1	Подзадача 2	Подзадача 3
Массив ТС (из перечня по проекту ТС)			A_{0j}			
Эксплуатационные данные	Параметры	безопасности	$A_{1j} A_{1j}^{\Delta}$			
		управления	$A_{2j} A_{2j}^{\Delta}$			
		аварийных систем	$A_{3j} A_{3j}^{\Delta}$			
		снабжения	A_{4j}			
Данные о деятельности персонала	Характеристики	осмотра	A_{5j}			
		технического обслуживания	A_{6j}			
		ремонта	A_{7j}			
		инспекций	A_{8j}			
цепочка «рейтинг 1» → «рейтинг 2» → «рейтинг 3»	Рейтинг первого уровня (рейтинг признаков)			G_{I1}	G_{I2}	G_{I3}
	↓			↓	↓	↓
	↓			$S_{G_{I1}}$	$S_{G_{I2}}$	$S_{G_{I3}}$
	↓	↓	↓	↓	↓	↓
	Рейтинг второго уровня (рейтинг задач)			$S_{i1}^{II} = \sum_l S_{G_{I1}}$	$S_{i2}^{II} = \sum_l S_{G_{I2}}$	$S_{i3}^{II} = \sum_l S_{G_{I3}}$
	↓	↓	↓	↓	↓	↓
	Рейтинг третьего уровня (рейтинг внимания)			для различных графов k_i для каждой отдельной задачи $S_i^{III} = \frac{S_i^{II}}{k_i}$		

Примечание: Представлена примерная схема построения кластеров для задач и возможных подзадач. Содержание признаков массивов может быть расширено и дополнено. В таблице предложена демонстрационная выборка для характеристик ТС. На практике могут быть случаи иных связей вершин кластеров. Для подзадачи 2 следует практиковать применение метрик и расстояний для анализа разброса и (или) отклонения количественных данных.

Таблица 6

Схема использования некоторых качественных методов анализа
после применения кластерного анализа безопасности ТС

Table 6

Scheme of the use of some methods of qualitative analysis after the application
of cluster analysis of TS safety

Условия применения <i>Terms of use</i>		Случаи <i>Cases</i>		
		S_{MIN}	< промежуток <	S_{MAX}
		Мнение эксперта в зависимости от замысла его задачи		
Для одной отдельной последовательности $S_{G_{II}}$	GAP-анализ	плохо (проблемно)	«серая» зона данных (требуется их уточнение)	хорошо
		хорошо		плохо (проблемно)
Для нескольких отдельных последовательно- стей S_{G_I}	SWOT-анализ	недостаток данных резерв данных для анализа		
	SNW-анализ	мало данных	резерв для рассмотрения	достаточно данных
		работа обеспече- на ресурсами		работа не обеспечена ресурсами

в пассивном (только анализировать), активном (предлагать решения и альтернативы) и комбинированном (получать/отдавать команды на уточнение данных и формировать оптимальное решение) режимах. Это позволяет получать более обоснованные управленческие решения, является средством повышения эффективности управления и проиллюстрировано для случаев и примеров, указанных в табл. 4–6.

При наличии недостатка данных для экспертного анализа безопасности ТС функционирование алгоритма для DATA MINING обеспечивает коммуникативные возможности для совместной деятельности многих специалистов (например, при надзоре или управлении), рассматривающих вопрос безопасности ТС с разных позиций. Работа таких алгоритмов базируется на использовании информации существующих баз данных для разных ТС, а также вносит в них новые комбинации данных. Способ DATA MINING здесь может применяться для получения решения в ходе синтеза знаний, сбора опыта анализа аналогичных задач из других сфер техники.

Оценка безопасности ТС осуществляется экспертом, так же как и генерация решений СППР, на каждом своем этапе по нескольким направлениям. Например, при определении условий, в которых требуется принятие решения; в проектном разрезе – для рассмотрения конкретного замысла или его альтернатив; при селекции в ходе выбора оптимального решения из множества рассмотренных или возможных альтернатив; в ходе внедрения, реализации или адаптации выбранного решения для конкретных условий и при выработке рекомендаций по безопасности ТС.

Исходные данные (входы СППР) представляют собой свойства, факторы или признаки (атрибуты), описывающие эксплуатационные или бизнес-процессы с ТС, для которых необхо-

димо принятие решения, а также знания специалистов или экспертов. Результатами исследования эксперта (выходами СППР) являются данные и (или) их анализ, на основе которых либо генерируются решения, либо они сами являются решениями.

Заключение

При использовании кластерного анализа исследователем проводится разбивка множества рассматриваемых объектов и признаков на удобные для понимания группы или кластеры. Однако с точки зрения эффективности дальнейших управленческих решений важно установить, кто определяет и проводит выборку признаков и их предварительную сортировку, т. е. до работы эксперта. Иначе подготовительная работа должна проводиться самими экспертом, специалистом по вопросам оценки безопасности ТС (внешним экспертом или исследователем в сфере эксплуатации ТС). Для решения задач классификации данных и выявления соответствующей структуры, создания комплексного документа для анализа со множеством признаков для использования методик кластерного подхода требуется реорганизация нормативных документов, где должно быть проведено четкое разграничение между требованиями к разработчикам как по обязательным мероприятиям, так и по дополнительным мероприятиям, которые проводятся в ходе экспертного рассмотрения проектов ТС.

Современные экспертные методы должны обладать свойствами аналитической обработки данных, возможностью построения математических моделей для программного анализа параметров. Хотя при решении ограниченных, целевых задач перебор параметров может проводиться самим экспертом вручную, но наличие математических моделей обработки данных и принятия решений позволяет автоматизировать анализ (вычислительными средствами), проводить перебор данных непрерывно и постоянно, без остановки, предлагая информацию для принятия управленческих решений. Для этого важно иметь в арсенале методы анализа для решения тактических и стратегических задач по оценке уровня безопасности, в том числе и с применением закономерностей теории графов.

Дополнительные данные и информация расширяют возможности и эффективность АСУТП и СППР. Это заключается в повышении обоснованности при принятии управленческих решений в условиях неопределенности либо недостатка данных или информации, быстрого изменения эксплуатационных условий.

Использование дополнительных методов анализа позволяет провести более глубокую и всестороннюю оценку уровня безопасности объектов и устройств. При проведении сравнения характеристик разных проектов ТС можно более детализировано выделить их преимущества и недостатки, «слабые» стороны [2].

Список литературы

1. Лобач Д. И., Ракитская Д. В. Сэйфеометрика. О количественной оценке величин при определении уровня безопасности // Промышленная безопасность. 2022. № 10. С. 46–47.
2. Лобач Д. И. О развитии экспертных возможностей для рассмотрения проектов оборудования и технологических решений // Системный анализ и прикладная информатика. 2023. № 2. С. 38–41. <https://doi.org/10.21122/2309-4923-2023-2-38-41>
3. Лобач Д. И. Новые проблемы, методология и возможности сэйфеометрики // Промышленная безопасность. 2023. № 1. С. 34–36.
4. Ковалев М. М. Образование для цифровой экономики // Цифровая трансформация. 2018. № 1 (2). С. 37–42.

5. **Фрич Р., Перегуд Е. Е., Мацевский С. В.** Избранные главы теории графов: Учебное пособие / Пер. с нем. Е. Е. Перегуда; Под ред. С. В. Мацевского. Калининград: Изд-во РГУ им. И. Канта, 2008. 204 с.
6. **Бериков В. С., Лбов Г. С.** Современные тенденции в кластерном анализе // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы». 2008. 26 с.
7. Operational Safety Performance Indicators for Nuclear Power Plants. International Atomic Energy Agency, TECDOC Series no. 1141. Vienna: IAEA, 2000.
8. **Шарафутдинов Р. Б., Кузнецов Л. А., Богданова Т. Ю.** Использование систем индикаторов безопасности зарубежными органами регулирования ядерной и радиационной безопасности // Ядерная и радиационная безопасность. 2008. № 2. С. 5–9.
9. **Хамаза А. А.** Предложения по внедрению риск-ориентированного подхода в контрольно-надзорной деятельности в области использования атомной энергии // Ядерная и радиационная безопасность. 2016. № 1. С. 1–6.
10. **Tryon R. C.** Cluster analysis. London: Ann Arbor Edwards Bros, 1939.
11. **Дюран Б., Оделл П.** Кластерный анализ. М.: Статистика, 1977. 128 с.
12. **Жамбю М.** Иерархический кластер-анализ и соответствия. М.: Финансы и статистика, 1988. 345 с.
13. **Мандель И. Д.** Кластерный анализ. М.: Финансы и статистика, 1988. 176 с.

References

1. **Lobach D. J., Rakitskaya D. V.** Safeometrics. About quantitative assessment of data for safety level determining. *Promyshlennaja bezopasnost* [*Promyshlennââ bezopasnost'*], 2022, no. 10, pp. 46–47. (in Russ.)
2. **Lobach D. J.** About development of expert possibilities for consideration of equipment projects and technological decisions. *Sistemny analiz i prikladnaja informatika* [*System analysis and applied information science*], 2023, no. 2, pp. 38–41. (in Russ.) <https://doi.org/10.21122/2309-4923-2023-2-38-41>
3. **Lobach D. J.** New problems, methodology and possibilities of Safeometrics. *Promyshlennaja bezopasnost* [*Promyshlennââ bezopasnost'*], 2023, no. 01, pp. 34–36. (in Russ.)
4. **Kovalev M. M.** Education for the Digital Economy. *Cifrovaja transformacija* [*Digital transformation*], 2018, no. 1 (2), pp. 37–42. (in Russ.)
5. **Frich R., Peregud E. E., Matsievsky S. V.** Selected chapters of graph theory: Textbook. Translated from German by E. E. Peregud; Edited by S. V. Matsievsky. Kaliningrad: Publishing House of the I. Kant Russian State University, 2008, 204 p. (in Russ.)
6. **Berikov V. S., Lbov G. S.** Modern trends in cluster analysis. *All-Russian competitive selection of review and analytical articles on the priority direction "Information and telecommunication systems"*, 2008, 26 p. (in Russ.)
7. Operational Safety Performance Indicators for Nuclear Power Plants. International Atomic Energy Agency, TECDOC Series no. 1141, Vienna: IAEA, 2000.
8. **Sharafutdinov R. B., Kuznetsov L. A., Bogdanova T. Yu.** The use of safety indicator systems by foreign regulatory bodies of nuclear and radiation safety. *Jadernaja i radiacionnaja bezopasnost* [*Nuclear and Radiation Safety*], 2008, no. 2, pp. 5–9. (in Russ.)
9. **Hamaza A. A.** Proposals for the introduction of a risk-based approach in control and supervisory activities in the field of atomic energy use. *Jadernaja i radiacionnaja bezopasnost* [*Nuclear and Radiation Safety*], 2016, no. 1, pp. 1–6. (in Russ.)
10. **Tryon R. C.** Cluster analysis. London, Ann Arbor Edwards Bros, 1939.
11. **Duran B., Odell P.** Cluster analysis. Moscow, Statistics publ., 1977, 128 p. (in Russ.)

12. **Zhambu M.** Hierarchical cluster-analysis and correspondence. Moscow, Finance and Statistics publ., 1988, 345 p. (in Russ.)
13. **Mandel I. D.** Cluster analysis. Moscow, Finance and Statistics publ., 1988, 176 p. (in Russ.)

Сведения об авторе

Лобач Дмитрий Иосифович, кандидат технических наук

Information about the Author

Dmitry J. Lobach, Candidate of Sciences (Technical Sciences)

*Статья поступила в редакцию 05.12.2023;
одобрена после рецензирования 09.04.2024; принята к публикации 09.04.2024*

*The article was submitted 05.12.2023;
approved after reviewing 09.04.2024; accepted for publication 09.04.2024*

Научная статья

УДК 004

DOI 10.25205/1818-7900-2024-22-1-49-61

Применение системы массового обслуживания для исследования характеристик канала связи в IoT-сетях

Сергей Валерьевич Малахов¹, Денис Олегович Якупов²
Ангелина Алексеевна Осипова³, Дарья Александровна Копылова⁴
Екатерина Александровна Зеленина⁵

Поволжский государственный университет телекоммуникаций и информатики,
Самара, Россия

¹s.malakhov@psuti.ru, <https://orcid.org/0009-0001-8666-6713>

²d.yakupov@psuti.ru, <https://orcid.org/00090009-0003-2371-0822>

³ange.osipowa2004@yandex.ru

⁴koroll.aleks@mail.ru

⁵0kate.zc0@gmail.com

Аннотация

Статья содержит подробную информацию о применении теории систем массового обслуживания (СМО) в сетях интернета вещей (IoT). Подробно рассматриваются математические модели, применяемые для анализа и оптимизации предоставления услуг в различных системах, включая IoT. Освещаются различные аспекты использования теории очередей в сетях IoT, такие как моделирование трафика, оптимизация передачи данных и использование стохастических моделей для более точного анализа. Исследование характеристик канала связи в сетях IoT представляет ключевую и актуальную проблему в условиях стремительного развития IoT-технологий. С ростом количества подключенных устройств становится критически важным обеспечить эффективность и надежность канала связи, а также оптимизировать использование ресурсов IoT-устройств. Данное исследование направлено на изучение и теоретический анализ характеристик канала связи в IoT с использованием систем массового обслуживания. В работе проводится анализ особенностей канала связи в IoT, изучаются методы моделирования канала, анализируются задержки передачи данных, оценивается и увеличивается пропускная способность, применяются методы систем массового обслуживания и исследуются области применения полученных результатов, прогнозируется развитие IoT, делается итоговый обзор научной работы. Интернет вещей (IoT) является сетью взаимодействующих устройств, использующих датчики и уникальные идентификаторы для обмена информацией. Широкое применение IoT в умных домах, энергетике, медицине, логистике и других секторах ускоряется благодаря современным технологиям искусственного интеллекта и машинного обучения.

Ключевые слова

сети IoT, теория очередей, системы массового обслуживания (СМО), моделирование трафика, оптимизация передачи данных, стохастические модели, высокая пропускная способность, низкая задержка, надежность

Для цитирования

Малахов С. В., Якупов Д. О., Осипова А. А., Копылова Д. А., Зеленина Е. А. Применение системы массового обслуживания для исследования характеристик канала связи в IoT-сетях // Вестник НГУ. Серия: Информационные технологии. 2024. Т. 22, № 1. С. 49–61. DOI 10.25205/1818-7900-2024-22-1-49-61

© Малахов С. В., Якупов Д. О., Осипова А. А., Копылова Д. А., Зеленина Е. А., 2024

The Use of a Queuing System to Study the Characteristics of the Communication Channel in IoT Networks

Sergey V. Malakhov¹, Denis O. Yakupov², Angelina A. Osipova³,
Daria A. Kopylova⁴, Ekaterina A. Zelenina⁵

Povolzhskiy State University of Telecommunications & Informatics,
Samara, Russian Federation

¹s.malakhov@psuti.ru, <https://orcid.org/0009-0001-8666-6713>

²d.yakupov@psuti.ru, <https://orcid.org/00090009-0003-2371-0822>

³ange.osipowa2004@yandex.ru

⁴koroll.aleks@mail.ru

⁵okate.ze0@gmail.com

Abstract

The article contains detailed information on the application of the theory of queuing systems (QMS) in the Internet of Things (IoT) networks. The article discusses in detail the mathematical models used to analyze and optimize the provision of services in various systems, including IoT. Various aspects of the use of queue theory in IoT networks are highlighted, such as traffic modeling, data transfer optimization, and the use of stochastic models for more accurate analysis. The study of the characteristics of the communication channel in IoT networks is a key and urgent problem in the context of the rapid development of IoT technologies. With the growing number of connected devices, it becomes critically important to ensure the efficiency and reliability of the communication channel, as well as optimize the use of IoT device resources. This study is aimed at studying and theoretical analysis of the characteristics of the communication channel in IoT using queuing systems. The paper analyzes the features of the communication channel in IoT, examines channel modeling methods, analyzes data transmission delays, evaluates and increases throughput, applies queuing system methods and explores the applications of the results obtained, predicts the development of IoT and makes a final review of scientific work. The Internet of Things (IoT) is a network of interacting devices that use sensors and unique identifiers to exchange information. The widespread use of IoT in smart homes, energy, medicine, logistics and other sectors is accelerating thanks to modern artificial intelligence and machine learning technologies.

Keywords

IoT networks, queueing theory, mass service systems (MSS), traffic modeling, data transfer optimization, stochastic models, high throughput, low latency, reliability

For citation

Malakhov S. V., Yakupov D. O., Osipova A. A., Kopylova D. A., Zelenina E. A. The use of a queuing system to study the characteristics of the communication channel in IoT networks. *Vestnik NSU. Series: Information Technologies*, 2024, vol. 22, no. 1, pp. 49–61 (in Russ.) DOI 10.25205/1818-7900-2024-22-1-49-61

Введение

Исследование характеристик канала связи в IoT-сетях является важной и актуальной проблемой в контексте быстрого развития IoT-технологий. С увеличением числа подключенных устройств становится критично не только обеспечение эффективности и надежности канала связи, но и оптимизация использования ресурсов IoT-устройств¹.

Устройства IoT полагаются на коммуникацию для взаимодействия друг с другом и внешними системами, такими как облачные платформы и мобильные приложения [1].

Цель данного исследования заключается в изучении и теоретическом анализе характеристик канала связи с использованием систем массового обслуживания в IoT.

Исследование включает анализ характеристик канала связи в IoT, определение основных концепций и технологий, изучение методов моделирования канала, анализ задержек передачи данных, оценку и увеличение пропускной способности, применение систем массового об-

¹ Characteristics of IoT. URL: <https://unstop.com/blog/characteristics-of-iot> (дата обращения: 19.02.2024).

служивания, исследование областей применения результатов, прогнозирование развития IoT и подведение итогов научной работы.

Основные понятия и технологии

Интернет вещей (IoT) представляет собой сеть устройств, которая взаимодействует друг с другом, используя датчики, аппаратное обеспечение и уникальные идентификаторы. Они позволяют устройствам обмениваться информацией между собой и передавать ее в центральную сеть, не требуя постоянного контроля со стороны человека².

Эта концепция применяется в разнообразных отраслях, начиная от мониторинга потребительского поведения в реальном времени и заканчивая оптимизацией рабочих процессов в различных секторах. IoT широко используется в умных домах, энергетике, медицине, логистике и многих других областях³.

Проникновение IoT в нашу повседневную жизнь ускорилося благодаря быстрому технологическому прогрессу. Современные разработки в области искусственного интеллекта, машинного обучения и передовых вычислений значительно повысили возможности IoT. Эти технологии помогают устройствам IoT извлекать ценные уроки из данных, предсказывать события и действовать автономно на основе накопленной информации. Например, с помощью искусственного интеллекта устройства IoT могут осуществлять анализ данных с датчиков для оптимизации энергопотребления в зданиях, снижения затрат и улучшения экологической эффективности.

Следует также отметить значительный вклад передовых вычислений, включая облачные и периферийные вычисления, в расширение возможностей IoT. Облачные вычисления предоставляют мощные ресурсы для хранения и обработки огромного объема данных, сгенерированных устройствами IoT, в то время как периферийные вычисления приближают вычислительные мощности к самим устройствам, обеспечивая оперативное принятие решений в реальном времени. Сила сочетания IoT, искусственного интеллекта и передовых вычислений привела к появлению «умных домов», «умных городов» и «умных производств», которые делают нашу жизнь удобнее и более экологически устойчивой⁴.

Системы массового обслуживания (СМО) – математические модели для анализа и оптимизации обслуживания в различных системах, включая IoT. СМО в IoT-сетях помогают аналитически моделировать передачу данных, повышая эффективность и качество обслуживания. Этот подход помогает оптимизировать трафик, улучшать процессы передачи данных и снижать нагрузку на сеть.

Основные понятия в теории СМО включают интенсивность входного потока, интенсивность обслуживания, размер очереди, время ожидания и время обслуживания. Эти концепции – основополагающие для анализа и оптимизации обслуживания в различных системах, включая IoT-сети.

Использование систем массового обслуживания в контексте интернета вещей представляет собой область исследования, требующую внимательного рассмотрения ряда дополнительных аспектов, влияющих на эффективность функционирования сети.

Первым из таких аспектов является моделирование трафика, что имеет важное значение в создании математических моделей, отражающих потоки данных в сети IoT. Этот процесс

² Why IoT Matters in Today's Modern World | Techreviewer. URL: <https://techreviewer.co/blog/why-iot-matters-in-todays-modern-world> (дата обращения: 19.02.24).

³ Что такое IoT и что о нем следует знать / Хабр. URL: <https://habr.com/ru/companies/otus/articles/549550/> (дата обращения: 19.02.24).

⁴ Интернет вещей (IoT): что это такое и в чем разница между обычным IoT и IIoT. URL: <https://cloud.vk.com/blog/iot-dlya-umnyh-chasov-iiot-umnyh-stankov-internet-veshchej> (дата обращения: 19.02.24).

обеспечивает возможность анализа и прогнозирования нагрузки на сеть, что является ключевым фактором для оптимизации ее работы.

Второй аспект эффективного управления передачей данных – это определение пропускной способности и задержек. Пропускная способность определяет максимальный объем данных, который может быть передан через сеть в единицу времени, тогда как задержка отражает время, необходимое для передачи данных от отправителя к получателю.

Третий аспект касается оптимизации передачи данных, что включает в себя настройку параметров сети для достижения максимальной эффективности и скорости передачи данных. Кроме того, улучшение обслуживания и снижение нагрузки направлены на повышение качества обслуживания конечных пользователей и обеспечение стабильной работы устройств IoT. Использование различных стратегий управления потоком данных, например, таких как FIFO или LIFO, где FIFO означает, что данные обрабатываются в порядке их поступления, а LIFO – что последние данные обрабатываются в первую очередь, представляет собой существенный аспект в обеспечении эффективной обработки данных в сети.

Немаловажным аспектом можно выделить анализ задержек передачи данных, а также аппроксимацию дисперсии задержек. Эти элементы являются важными для понимания изменчивости времени передачи данных в сети и принятия соответствующих мер по оптимизации работы системы⁵.

Применение СМО в сетях передачи данных может быть представлено следующими примерами.

1. Гибкий коммутатор (Softswitch): использование СМО для обработки обычных телефонных звонков и звонков на бесплатные номера «800». В таких случаях возможно возвращение части заявок на обслуживание на вход коммутатора, что создает дополнительный поток входящих заявок (поток обратной связи).

2. Моделирование работы центра коммутации мобильной связи (MSC): исследование взаимодействия MSC с базами данных регистра местоположения (HLR), регистра перемещений (VLR), центра аутентификации (AUC) и другими компонентами сети для оптимизации процессов передачи данных.

3. Двухфазная бесконечно линейная система: использование математической модели системы массового обслуживания требований случайного объема для исследования двухфазной бесконечно линейной системы в сети.

4. Расчет параметров линий связи информационно-управляющих систем: применение моделей СМО для решения задач, связанных с оптимизацией параметров линий связи в информационно-управляющих системах.

5. Исследование процессов передачи видеопотоков в пакетных сетях: применение СМО для анализа и оптимизации процессов передачи видеопотоков в сетях передачи данных, что позволяет улучшить качество и эффективность передачи мультимедийного контента.

6. Управление качеством обслуживания (QoS): использование СМО для управления качеством обслуживания в сетях с высокой пропускной способностью, таких как сети 5G и выше. Это позволяет оптимизировать передачу данных, учитывая требования пользователей и ограничения сети.

7. Распределение ресурсов в сетях с облачным хостингом: использование СМО для распределения ресурсов в сетях с облачным хостингом, включая выделение виртуальных машин и ресурсов хранения. Это позволяет оптимизировать использование ресурсов и уменьшить задержки в передаче данных.

8. Моделирование работы сети IoT: использование СМО для моделирования работы сети IoT, включая анализ трафика, задержек и потерь пакетов. Это позволяет оптимизировать работу сети и улучшить качество обслуживания пользователей.

⁵ Основы систем массового обслуживания: определение, компоненты и методы анализа. URL: <https://nauchniestati.ru/spravka/sistemy-massovogo-obsluzhivaniya/> (дата обращения: 19.02.24).

Моделирование канала связи в IoT

СМО позволяют аналитически моделировать и оптимизировать процессы передачи данных, улучшая их эффективность и качество обслуживания. Это важно для обеспечения работы IoT-сетей, которые требуют высокой пропускной способности, низких задержек и надежности [2; 3].

Метод моделирования канала связи в IoT-сетях включает следующие этапы.

1. Определение характеристик канала связи – пропускная способность, задержки, потери данных.
2. Выбор модели СМО, в зависимости от требований.
3. Анализ трафика – интенсивность входного потока и обслуживания.
4. Очередь и время ожидания – размер и время ожидания заявок в системе.
5. Анализ времени обслуживания и оптимизация – время обслуживания и улучшение процессов передачи данных [4].
6. Оптимизация качества обслуживания и нагрузки на сеть – улучшение времени ожидания, времени обслуживания и пропускной способности, а также снижение нагрузки на сеть.

Использование стохастических моделей для моделирования канала связи в IoT позволяет учесть случайные переменные и вероятностные законы. Стохастические модели широко применяются в различных областях, включая информатику, физику, экономику и другие, и позволяют более точно учитывать случайные факторы в моделировании процессов обслуживания⁶.

Моделирование каналов передачи данных в IoT является важным аспектом исследования IoT-сетей. Оно позволяет анализировать и оптимизировать процессы передачи данных, улучшая качество обслуживания и эффективность сети. Существует множество протоколов беспроводной сети IoT, каждый с особенностями и недостатками, выбор зависит от требований. Для моделирования используются различные методы, включая стохастические модели и моделирование систем NB-IoT. Стохастические модели учитывают случайные факторы, обеспечивая точное моделирование процессов передачи данных и обслуживания. Одним из ключевых аспектов является выбор подходящего протокола и оптимизация процессов передачи данных, что существенно влияет на работоспособность IoT-сетей⁷.

Системы массового обслуживания являются эффективным инструментом для моделирования канала связи в IoT-сетях. Они позволяют аналитически моделировать и оптимизировать процессы передачи данных, улучшая их качество обслуживания и используя для моделирования различных аспектов, а также для оценки характеристик канала связи в IoT-сетях.

Одним из примеров применения СМО для оценки характеристик канала связи является моделирование мультиплексирования разноприоритетных сообщений. Системы массового обслуживания позволяют распределять каналы передачи за потоками данных с разными приоритетами, что важно для обеспечения успешной работы IoT-сетей. Кроме того, СМО могут использоваться для расчета характеристик линий связи информационно-управляющих систем.

Выбор варианта организации системы массового обслуживания, обслуживающей неоднородные заявки в зависимости от стоимости содержания каналов также является важным аспектом использования СМО для оценки характеристик канала связи в IoT-сетях. Они позволяют выбирать оптимальную организацию обслуживания в зависимости от стоимости содержания каналов и также могут использоваться для моделирования различных типов каналов

⁶ Ограничение пропускной способности: что это такое и как оно влияет на работу сетей. URL: <https://obzorposudy.ru/polezno/ogranichenie-propusknoi-sposobnosti-ponyatie-i-primery> (дата обращения: 19.02.24).

⁷ Сравнительный анализ стандартов связи для сетей IoT / Хабр. URL: <https://habr.com/ru/companies/msw/articles/720518/> (дата обращения: 19.02.24).

передачи данных в IoT-сетях, таких как беспроводные каналы связи, кабельные каналы связи и другие^{8,9}.

Характеристики канала связи в IoT

Задержки в передаче данных влияют на производительность IoT. Они могут привести к сбоям в критически важных функциях, требующих реального времени. Перегрузка сети, расстояние и обработка данных влияют на задержки [5]. Важно поддерживать задержки для обеспечения качества обслуживания.

Технические задержки связаны с ограничениями инфраструктуры, пропускной способностью и временем передачи данных. Временные задержки происходят из обработки данных на устройствах, включая сервера и конечные устройства¹⁰. Оптимизация обработки данных и производительности сетей сокращает временные задержки.

Расстояние между устройствами, скорость передачи данных, количество устройств в сети и технические характеристики сети – ключевые факторы, влияющие на задержку передачи данных. Большие расстояния и низкая скорость увеличивают временные задержки. Увеличение числа устройств может вызвать узкие места и конфликты, приводя к задержкам. Технические параметры сети, такие как пропускная способность и качество соединения, также влияют на скорость и надежность передачи данных.

Оценка и оптимизация пропускной способности в IoT критичны для эффективной передачи данных. NB-IoT, обладающий низкой скоростью передачи, подходит для устройств с ограниченной мощностью и долгим сроком службы батареи, характерен для стационарных устройств с небольшим объемом данных.

Для улучшения оценки канала в NB-IoT исследователи предложили методы, включая перекрывающийся полосами SVD LMMSE, чтобы повысить точность и уменьшить сложность оценки канала.

Для повышения пропускной способности в сетях IoT применяют различные методы. Например, расширение пропускной способности NB-IoT позволяет использовать внутридиапазонное соединение, даже если ширина канала составляет всего 200 кГц [6].

Также важно использовать многопередатчиковые системы, которые увеличивают доступную пропускную способность. Для балансировки нагрузки и увеличения пропускной способности подключенных устройств целесообразно использовать сети IoT без ячеек [7].

Потеря пакетов – это ситуация, когда данные не достигают получателя из-за различных причин. В IoT-сетях проблемы с потерей пакетов возникают из-за перегрузки сети, слабого сигнала, помех, системного шума, несовершенных настроек и старого оборудования [8].

Беспроводные сети, особенно Wi-Fi, более подвержены потере пакетов, чем проводные соединения. Для улучшения производительности и надежности сети IoT важно анализировать и исправлять эти причины [9].

Увеличение пропускной способности, глубокая инспекция пакетов и обновление ПО и оборудования помогают предотвратить потерю пакетов. Использование проводных соединений, мониторинг производительности и оптимизация MTU и буфера на маршрутизаторе снижают вероятность потерь. VPN и замена неисправного оборудования также важны для надежности сети.

⁸ Системы тестирования сетей 5G. URL: <https://2test.ru/news/publikatsii/sistemy-testirovaniya-setey-5g/> (дата обращения: 19.02.24).

⁹ Погружение в мир AI: курсы, проекты, советы. URL: https://aicomb.ru/primenenie-ii-v-povsednevnoj-zhizni/intellektualnye_sistemy_v_telekommunikatsijah_uluchshenie_kachestva_svjazi_i_optimizatsija_setej/ (дата обращения: 19.02.24).

¹⁰ What is Packet Loss. URL: <https://www.fortinet.com/it/resources/cyberglossary/what-is-packet-loss> (дата обращения: 19.02.2024).

Теоретические модели и анализ

Существует разнообразие моделей систем массового обслуживания, применяемых для изучения и улучшения работы систем.

1. Модель М/М/1:

- М означает экспоненциальный (пуассоновский) поток клиентов;
- М означает экспоненциальное время обслуживания;
- 1 означает одно обслуживающее устройство.

Эта модель описывает систему, где поступают клиенты в соответствии с экспоненциальным распределением (М), время обслуживания также подчиняется экспоненциальному распределению (М), и имеется только один сервер. Она широко применяется для анализа производительности систем, где не существует ограничений на длину очереди и заявки поступают в соответствии с процессом Пуассона.

2. Модель М/М/с:

- М означает экспоненциальный поток клиентов;
- М означает экспоненциальное время обслуживания;
- с означает число параллельных обслуживающих устройств.

Эта модель является расширением предыдущей модели и включает несколько параллельных серверов. Применяется в случаях, когда есть несколько серверов, которые могут одновременно обслуживать клиентов. Оптимизация производительности системы в этой модели происходит путем выбора оптимального количества обслуживающих устройств.

3. Модель М/Г/1:

- М означает экспоненциальный поток клиентов;
- Г означает произвольное (общее) время обслуживания;
- 1 означает одно обслуживающее устройство.

Эта модель представляет собой расширение модели М/М/1, где время обработки заявок уже необязательно подчиняется экспоненциальному распределению, а может быть произвольным (обозначается как Г). Это делает модель более гибкой при анализе реальных сценариев, где время обработки может варьироваться в зависимости от различных факторов.

4. Модель М/Г/с:

- М означает экспоненциальный поток клиентов;
- Г означает произвольное время обслуживания;
- с означает число параллельных обслуживающих устройств.

Аналогично модели М/Г/1, здесь учитывается возможность произвольного времени обработки заявок, а также присутствует несколько параллельно работающих обслуживающих устройств. Оптимальное количество устройств подбирается для максимизации производительности системы, учитывая различные характеристики времени обработки и потребности в обслуживании.

Каждая из этих моделей систем массового обслуживания предоставляет математический инструментарий для анализа и оптимизации систем массового обслуживания с различными характеристиками поступления и обработки заявок¹¹.

При оценке эффективности систем массового обслуживания (СМО) важно учитывать несколько ключевых показателей.

1. Среднее время пребывания заявки в СМО: показывает время, проведенное заявкой в системе до обслуживания.

2. Вероятность обслуживания заявки: определяет вероятность успешного обслуживания заявки.

¹¹ Модель системы массового обслуживания: основные принципы и применение в реальной жизни. URL: <https://nauchniestati.ru/spravka/model-sistemy-massovogo-obsluzhivaniya/> (дата обращения: 19.02.24).

3. Среднее занятое количество каналов: низкое значение может указывать на неэффективное использование ресурсов.

4. Среднее количество занятых мест в очереди: высокое число может привести к отказам.

Анализируя ключевые параметры работы систем массового обслуживания в моделях М/М/1, М/М/с, М/Г/1 и М/Г/с, можно выделить следующие характеристики (табл. 1).

Таблица 1

Анализ работы систем массового обслуживания
в моделях М/М/1, М/М/с, М/Г/1 и М/Г/с

Table 1

Analysis of the operation of queuing systems
in the models М/М/1, М/М/с, М/Г/1 and М/Г/с

Параметр	М/М/1	М/М/с	М/Г/1	М/Г/с
Среднее время пребывания заявки в СМО	Низкое	Зависит от (с)	Разнообразие во времени обслуживания может привести к изменчивости времени пребывания заявки в системе	Зависит от (с)
Вероятность обслуживания заявки	Постоянная, не зависит от (с)	Зависит от (с)	Гибкая, учитывает различные распределения времени обслуживания	Пропорциональная (с)
Среднее занятое количество каналов	1	С	Гибкое	С
Среднее количество занятых мест в очереди	Заявки мгновенно обслуживаются	Зависит от интенсивности поступления заявок и количества обслуживающих устройств	Может меняться в зависимости от распределения	Значение зависит от интенсивности поступления заявок, количества обслуживающих устройств и характеристик распределения времени обслуживания

Анализ параметров позволяет понять, как различные модели влияют на время пребывания заявок, вероятность их обслуживания, использование каналов связи и эффективность управления очередями. Выбор оптимальной модели зависит от конкретных требований системы и баланса между ресурсами и производительностью. Важно понимать, какие параметры влияют на эффективность таких каналов.

1. Пропускная способность канала: определяет объем данных, передаваемых за единицу времени.

2. Задержка передачи данных: время от источника к приемнику. Минимизация повышает отзывчивость и скорость передачи.
3. Джиттер: изменение задержки передачи данных.
4. Пропускная способность очереди: максимальный объем данных в очереди.
5. Ошибки передачи данных: могут снизить производительность и вызывать повторную передачу.
6. Протокол передачи данных: выбор протокола влияет на скорость, надежность и стабильность связи¹².

В табл. 2 приведен анализ параметров разных моделей систем массового обслуживания.

Таблица 2

Анализ параметров разных моделей систем массового обслуживания

Table 1

Analysis of parameters of different models of queuing systems

Параметр	М/М/1	М/М/с	М/Г/1	М/Г/с
Пропускная способность канала	Одна единица (примерно 1 Gbps)	В зависимости от (с)	Гибкая	Пропорционально (с)
Задержка передачи данных	Низкая	Варьируется в зависимости от (с)	Может варьироваться из-за неоднородности времени обслуживания	Варьируется в зависимости от (с)
Джиттер	Низкий	Варьируется от модели к модели	Может проявляться в неоднородности времени ожидания и обработки заявок	Варьируется от модели к модели
Пропускная способность очереди	Обслуживается только одна заявка за раз и не возникает очереди	Пропускная способность очереди обычно не является определяющим параметром	Может быть ограничена временем обработки одной заявки и интенсивностью поступления новых заявок	Зависима от количества (с), характеристик времени обработки заявок и наличия ограничений на длину очереди
Ошибка передачи данных	Заявки обслуживаются мгновенно без очереди, вероятность возникновения ошибок передачи данных низкая	При увеличении числа устройств в системе может уменьшаться вероятность ошибок из-за распределения нагрузки	При возможном изменчивом времени обслуживания заявок возможно увеличение вероятности ошибок передачи данных	Может изменяться в зависимости от (с) и характеристик времени обработки заявок
Тип используемого протокола	В основном TCP	Использование различных протоколов в зависимости от специфики обслуживаемых устройств и потребностей сети	Использование различных протоколов в зависимости от конкретных условий, требований системы и возможностей оборудования	Варьируется в зависимости от числа обслуживаемых устройств и требований конкретной ситуации

¹² Компьютерное моделирование. СМО. URL: https://ami.nstu.ru/~gulyaeva/cm/materials/lec1_cm.pdf/ (дата обращения: 19.02.24).

Анализируя эти параметры, можно определить различия между моделями систем массового обслуживания в контексте различных параметров, которые влияют на их эффективность и производительность [10; 11].

Увеличение числа каналов в системе массового обслуживания улучшает производительность и обработку заявок, снижая время ожидания и повышая пропускную способность. Равномерное распределение нагрузки между каналами через оптимизацию позволяет эффективнее использовать ресурсы и предотвращать перегрузки. Оптимальная стратегия должна сбалансировать эффективность обработки и удовлетворенность клиентов [12].

Применение и перспективы

Рассмотрим прогнозирование дальнейшего развития IoT-сетей на основе исследования, проведенного крупнейшим российским оператором связи МТС: к концу 2025 г. около 25 % всех IoT-устройств в России будут подключены через eSIM. Прогнозируются среднегодовые темпы роста рынка IoT в России на уровне 12 % после спада в текущем году. К 2026 г. рынок может достичь объема в 208,5 млрд рублей. Отрасли энергетики, ЖКХ, недвижимости и транспорта останутся в лидирующей позиции по внедрению IoT. К 2025 г. общее количество интеллектуальных счетчиков коммунальных ресурсов в России увеличится до 39 млн. Прогнозируется утроение количества «умных» колонок и почти пятикратный рост устройств умного дома к 2025 г.¹³

Прогнозы показывают ожидаемое развитие сферы IoT в различных частях мира и раскрывают стратегические намерения региональных лидеров в области инноваций и цифровизации.

В США: по прогнозам IDC, к 2023 году глобальный рынок IoT вырастет до 1,1 триллиона долларов, где США будут лидировать по объему IoT-инвестиций. В Китае: Counterpoint Research ожидает, что к 2025 г. Китай станет крупнейшим рынком IoT смарт-устройств в мире. В Европейском союзе: Евросоюз планирует инвестировать более €20 млрд в развитие IoT к 2027 г., что приведет к значительному расширению IoT-сетей в странах ЕС.

Анализ прогнозов указывает на активное развитие IoT-сетей как в России, так и в других странах мира. Увеличение числа подключенных устройств, расширение инфраструктуры и инвестиции в инновации говорят о перспективности этой отрасли в будущем.

Заключение

Применение теории очередей к каналам связи в сетях IoT является перспективной областью исследований, которая может помочь решить ряд актуальных проблем. Быстрый рост устройств IoT приводит к увеличению сетевой нагрузки, что может приводить к задержкам, потере пакетов и снижению надежности сети. Система массового обслуживания может быть использована для анализа поведения этих сетей и разработки более эффективных протоколов связи, которые могут помочь уменьшить эти проблемы. Некоторые из последних технологических трендов, применяемых в исследовании каналов связи:

- 1) исследование с использованием 5G-технологий;
- 2) использование искусственного интеллекта (AI) в анализе каналов связи;
- 3) исследование ультраширокополосных (UWB) технологий;
- 4) применение программно-определяемых сетей (SDN) и сетей функций (NFV).

Однако другая ключевая проблема в сетях IoT заключается в обеспечении справедливости и качества обслуживания (QoS) для разных типов трафика. Например, приложения, которым

¹³ Тренды IoT: что ждет интернет вещей в ближайшие годы. URL: <https://vc.ru/mts/456729-trendy-iot-chto-zhdet-internet-veshchey-v-blizhayshie-gody/> (дата обращения: 19.02.24).

нужно быстро передавать данные почти сразу, например видео- или голосовые звонки, требуют короткой задержки и высокой скорости передачи данных. Но, напротив, приложениям, работающим с данными, передаваемыми не сразу, например отправка файлов или электронных писем, время не так критично. Система массового обслуживания может быть использована для разработки алгоритмов планирования, которые приоритизируют разные типы трафика на основе их требований QoS, обеспечивая доступность необходимых ресурсов для всех устройств и приложений.

Наконец, использование алгоритмов машинного обучения в сетях IoT может приводить к увеличению вычислительной сложности и требованиям к ресурсам. Система массового обслуживания может быть использована для анализа производительности этих алгоритмов и разработки более эффективных стратегий распределения ресурсов, которые могут помочь уменьшить энергопотребление и улучшить производительность сети.

В целом применение системы массового обслуживания к каналам связи в сетях IoT может помочь решить ряд ключевых проблем, включая сетевую нагрузку, обеспечение QoS и распределение ресурсов. Используя принципы системы массового обслуживания, исследователи могут разработать более эффективные протоколы связи, алгоритмы планирования и стратегии распределения ресурсов, которые могут помочь улучшить производительность и надежность сетей IoT.

Список литературы

1. **Ядровская М. В., Поркшеян М. В., Синельников А. А.** Перспективы технологии интернета вещей // *Advanced Engineering Research*. 2021. Т. 21, № 2. С. 207–217. DOI: <https://doi.org/10.23947/2687-1653-2021-21-2-207-217>
2. Методы принятия управленческих решений: Учебное пособие / Под ред. П. В. Иванова. Ростов н/Д: Феникс, 2014. 413 с.
3. **Аль-Днебат Сайд Али.** Применение сетей массового обслуживания для исследования процессов передачи видеопотоков в пакетных сетях: Автореф. ... дис. канд. техн. наук: 05.12.13 / науч. рук. А. Ф. Ярославцев. Новосибирск, 2004. 17 с.
4. **Петухов О. А., Морозов А. В., Петухова Е. О.** Моделирование: системное, имитационное, аналитическое: Учеб. пособие. 2-е изд., испр. и доп. СПб.: Изд-во СЗТУ, 2008. 288 с.
5. **Chen Y, Lu L, Yu X, Li X.** Adaptive Method for Packet Loss Types in IoT: An Naive Bayes Distinguisher // *Electronics*. 2019. Vol. 8(2). P. 134. DOI: <https://doi.org/10.3390/electronics8020134>
6. **Shi V. T., Nhg D. R.** Channel Estimation Optimization Model in Internet of Things based on MIMO/OFDM with Deep Extended Kalman Filter // *Advances in Engineering and Intelligence Systems*. 2022. Vol. 001(02). P. 822. DOI: 10.22034/aeis.2022.341792.1020
7. **Jewel M. K. H., Zakariyya R. S., Lin F.** On Channel Estimation in LTE-Based Downlink Narrowband Internet of Things Systems // *Electronics*. 2021. Vol. 10(11). P. 1246. DOI: <https://doi.org/10.3390/electronics10111246>
8. **Jung J.-Y., Lee J.-R.** Throughput and Packet Loss Probability Analysis of Long Range Wide Area Network // *Applied Sciences*. 2021. Vol. 11(17). P. 8091. DOI: <https://doi.org/10.3390/app11178091>
9. **Abbas, Qamar & Hassan, Syed & Pervaiz, Haris & Ni, Qiang.** A Markovian Model for the Analysis of Age of Information in IoT Networks. *IEEE Wireless Communication Letters*. 2021. DOI: 10.1109/LWC.2021.3075160.
10. **Захарикова Е. Б.** Математические модели сетей массового обслуживания в виде «вход-состояние-выход» // *Экономика, статистика и информатика. Вестник УМО*. 2012. № 6. С. 188192.

11. **Емельянов А. А.** Модели процессов массового обслуживания // Прикладная информатика. 2008. № 5(17). С. 92130.
12. **Chen, Kevin C. W. and Wei, Kuo-Chiang (John) and Chen, Zhihong** Disclosure, Corporate Governance, and the Cost of Equity Capital: Evidence from Asia's Emerging Markets (June 2003). DOI: <http://dx.doi.org/10.2139/ssrn.422000>

References

1. **Yadrovskaya M. V., Porksheyev M. V., Sinelnikov A. A.** Prospects for Internet of Things technology. *Advanced Engineering Research*, 2021, vol. 21, no. 2, pp. 207–217. DOI: <https://doi.org/10.23947/2687-1653-2021-21-2-207-217> (in Russ.)
2. Methods for making management decisions: textbook. Ivanova P. V. (ed.). Rostov na Done, Phoenix publ., 2014, 413 p. (in Russ.)
3. Why IoT Matters in Today's Modern World | Techreviewer. URL: <https://techreviewer.co/blog/why-iot-matters-in-todays-modern-world> (date of application: 02.19.24).
4. **Al-Dnebat Said Ali** Application of queuing networks to study the processes of transmitting video streams in packet networks: abstract of thesis. Dis. Ph.D. Novosibirsk, 2004, 17 p. (in Russ.)
5. **Chen Y, Lu L, Yu X, Li X.** Adaptive Method for Packet Loss Types in IoT: An Naive Bayes Distinguisher. *Electronics*, 2019, vol. 8(2), p. 134. DOI: <https://doi.org/10.3390/electronics8020134>
6. **Shi V. T., Nhg D. R.** Channel Estimation Optimization Model in Internet of Things based on MIMO/OFDM with Deep Extended Kalman Filter. *Advances in Engineering and Intelligence Systems*, 2022, vol. 001(02), pp. 822. DOI: 10.22034/aeis.2022.341792.1020
7. **Jewel M. K. H., Zakariyya R. S., Lin F.** On Channel Estimation in LTE-Based Downlink Narrowband Internet of Things Systems. *Electronics*, 2021, vol. 10(11), p. 1246. DOI: <https://doi.org/10.3390/electronics10111246>
8. **Jung J-Y, Lee J-R.** Throughput and Packet Loss Probability Analysis of Long Range Wide Area Network. *Applied Sciences*, 2021, vol. 11(17), p. 8091. DOI: <https://doi.org/10.3390/app11178091>
9. **Abbas, Qamar & Hassan, Syed & Pervaiz, Haris & Ni, Qiang.** A Markovian Model for the Analysis of Age of Information in IoT Networks. *IEEE Wireless Communication Letters*, 2021. DOI: 10.1109/LWC.2021.3075160.
10. **Zakharikova E. B.** Mathematical models of queuing networks in the form of “input-state-output”. *Economics, statistics and computer science. UMO Bulletin*, 2012, no. 6, pp. 188192. (in Russ.)
11. **Emelyanov A. A.** Models of queuing processes. *Applied informatics*, 2008, no. 5(17), pp. 92130. (in Russ.)
12. **Chen, Kevin C. W. and Wei, Kuo-Chiang (John) and Chen, Zhihong.** Disclosure, Corporate Governance, and the Cost of Equity Capital: Evidence from Asia's Emerging Markets (June 2003). DOI: <http://dx.doi.org/10.2139/ssrn.422000>

Сведения об авторах

Малахов Сергей Валерьевич, кандидат технических наук

SPIN-код: 1089-9215

AuthorID: 926021

Якупов Денис Олегович, аспирант

SPIN-код: 8692-1106

AuthorID: 1175874

Осипова Ангелина Алексеевна, студентка

Копылова Дарья Александровна, студентка

Зеленина Екатерина Александровна, студентка

Information about the Authors

Sergey V. Malakhov, Candidate of Technical Sciences

SPIN code: 1089-9215

Author ID: 926021

Denis O. Yakupov, Post-Graduate Student

SPIN code: 8692-1106

Author ID: 1175874

Angelina A. Osipova, Student

Daria A. Kopylova, Student

Ekaterina A. Zelenina, Student

*Статья поступила в редакцию 14.03.2024;
одобрена после рецензирования 11.04.2024; принята к публикации 11.04.2024*

*The article was submitted 14.03.2024;
approved after reviewing 11.04.2024; accepted for publication 11.04.2024*

Научная статья

УДК 519.6

DOI 10.25205/1818-7900-2024-22-1-62-72

Оценка приземной концентрации метана на основе алгоритма ансамблевого фильтра Калмана с привлечением транспортной химической модели

Марина Владимировна Платонова¹

Василий Дмитриевич Котлер²

Екатерина Георгиевна Климова³

Федеральный исследовательский центр информационных и вычислительных технологий,
Новосибирск, Россия

¹gumoznaya@gmail.com, <https://orcid.org/0000-0003-4359-8922>

Аннотация

В статье приводятся результаты численных экспериментов с модельными данными по оценке приземной концентрации метана на всей территории Земли. Имитация данных наблюдений для модельных экспериментов производилась по прогнозам модели переноса и диффузии MOZART-4. Приземная концентрация является значением, задаваемым на нижней границе модели. Ее изменение по времени характеризует эмиссию метана с поверхности Земли. Численные эксперименты с модельными данными проведены как с имитацией сети наземных измерений, так и с имитацией спутниковых наблюдений. Алгоритм поиска оценки приземной концентрации основан на использовании ансамблевого фильтра Калмана. Показано, что разработанный алгоритм позволяет оценивать значения приземной концентрации метана по данным наблюдений. Точность оценки зависит от распределения в пространстве и времени используемых данных измерений.

Ключевые слова

усвоение данных, оценка приземной концентрации метана, модельные эксперименты

Финансирование

Работа выполнена в рамках государственного задания Минобрнауки России для Федерального исследовательского центра информационных и вычислительных технологий.

Благодарности

Авторы статьи выражают глубокую благодарность коллегам профессору Анатолию Алексеевичу Лагутину и доценту Егору Юрьевичу Мордвину за предоставленные данные и полезные дискуссии.

Для цитирования

Платонова М. В., Котлер В. Д., Климова Е. Г. Оценка приземной концентрации метана на основе алгоритма ансамблевого фильтра Калмана с привлечением транспортной химической модели // Вестник НГУ. Серия: Информационные технологии. 2024. Т. 22, № 1. С. 62–72. DOI 10.25205/1818-7900-2024-22-1-62-72

© Платонова М. В., Котлер В. Д., Климова Е. Г., 2024

Estimation of Surface Methane Concentration based on the Ensemble Kalman Filter Algorithm using a Transport Chemical Model

Marina V. Platonova¹, Vasily D. Kotler²
Ekaterina G. Klimova³

Federal Research Center for Information and Computational Technologies,
Novosibirsk, Russian Federation

¹gumoznaya@gmail.com, <https://orcid.org/0000-0003-4359-8922>

Abstract

The article presents the results of numerical experiments using model data to estimate ground-level methane concentrations using the MOZART-4 model. Various approaches to integrating observational data and their application to various scientific and practical applications are discussed, including monitoring and analysis of methane sources, both anthropogenic and natural. These results illustrate the practical use of data assimilation to collect statistical data on the dynamics of emissions activity in specific subregions, which can be useful for estimating activity levels and processing large data sets to identify the most interesting and potentially promising areas for obtaining more detailed data analysis.

Keywords

data assimilation, estimation of surface methane concentration, model experiments

Funding

The work was carried out within the framework of the state assignment of the Russian Ministry of Education and Science for the Federal Research Center for Information and Computing Technologies.

Acknowledgements

The authors of the article express their gratitude to their colleagues Professor Anatoly Alekseevich Lagutin and Associate Professor Egor Yurievich Mordvin for the data provided and useful discussions.

For citation

Platonova M. V., Kotler V. D., Klimova E. G. Estimation of surface methane concentration based on the ensemble Kalman filter algorithm using a transport chemical model. *Vestnik NSU. Series: Information Technologies*, 2024, vol. 22, no. 1, pp. 62–72 (in Russ.) DOI 10.25205/1818-7900-2024-22-1-62-72

Введение

Актуальность научных исследований в области мониторинга парниковых газов неоспорима. Метан является одним из основных парниковых газов и играет ключевую роль в глобальном потеплении и изменении климата. Его влияние на окружающую среду, как в атмосфере, так и на поверхности Земли, является предметом широкого научного интереса и обсуждений. Одним из современных методов, активно применяемых в области моделирования процессов в окружающей среде, является метод обратного моделирования. Методы обратного моделирования представляют собой совместное использование данных наблюдений и математической модели для получения оценки заданных величин. Усвоение данных представляет собой частный случай обратного моделирования, который включает последовательное получение оценок по времени на основе данных наблюдений [1].

Существуют два основных подхода к решению задачи усвоения данных: вариационный и динамико-стохастический. Вариационные методы базируются на поиске минимума функционала, описывающего отклонение модельного решения и данных измерений в заданной метрике [2; 3]. Динамико-стохастический подход основан на применении теории оценивания, в частности, фильтра Калмана. Реализация алгоритмов усвоения данных, основанных на фильтре Калмана, в задачах моделирования окружающей среды является сложной вычислительной задачей из-за высокой размерности используемых в алгоритме векторов и матриц. По этой

причине в настоящее время используются ансамблевые методы, такие как ансамблевый фильтр Калмана и ансамблевое сглаживание Калмана [4–12]. Использование данных спутникового наблюдения в системах усвоения данных обеспечивает широкий географический охват и непрерывный мониторинг состояния окружающей среды, в то время как результаты расчета моделей позволяют учитывать сложные физические и химические процессы, происходящие в атмосфере [13; 14]. Комбинируя эти источники информации, можно достичь получения оптимальной оценки концентраций парникового газа.

Современные методы оценки концентрации и эмиссии парниковых газов используют данные наблюдений, включая спутниковые, с высоким разрешением [15]. Эти методы основаны на усвоении данных и обратном моделировании. Однако такие методы требуют значительных вычислительных ресурсов, альтернативным подходом является использование ансамблевых методов, таких как ансамблевый фильтр Калмана и ансамблевое сглаживание Калмана [4–12; 16]. Эти методы являются технически более простыми в реализации и могут быть применены последовательно по времени. Большое количество работ посвящено оценке эмиссии парниковых газов на основе ансамблевого фильтра Калмана. Отметим работу [16], в которой приводятся результаты численных экспериментов с модельными данными по оценке эмиссии метана, при этом применяется алгоритм ансамблевого фильтра Калмана LETKF.

В настоящей работе представлены результаты численных экспериментов с модельными данными по оценке приземной концентрации метана с использованием транспортной химической модели MOZART-4 [4–7; 11–14]. Приземная концентрация метана является нижним граничным условием модели и ее изменчивость по времени характеризует динамику эмиссии метана с поверхности Земли. В работе используется алгоритм ансамблевого фильтра Калмана LETKF [16]. Описана реализация алгоритма для оценки приземной концентрации метана на поверхности Земного шара. Приводятся результаты численных экспериментов с модельными данными, при этом рассматриваются различные варианты задания модельных данных. Численные эксперименты с модельными данными являются важным этапом при разработке системы усвоения данных, они позволяют оценить эффективность системы усвоения данных и возможность оценивать требуемый параметр с заданной точностью.

Алгоритм ансамблевого фильтра Калмана

В данной работе применяется динамико-стохастический подход к задаче усвоения данных, который основан на использовании ансамблевого фильтра Калмана [17]. Ниже приводится описание используемого в численных экспериментах алгоритма ансамблевого фильтра Калмана LETKF.

Алгоритм усвоения данных применяется для оценки приземной концентрации метана и состоит из шага прогноза и шага анализа. Шаг анализа в общем случае имеет вид [4–7; 11; 12; 17]:

$$x_k^a = x_k^f + K_k[y_0^k - H(x_k^f)].$$

В этих формулах k – номер шага по времени, K_k – весовая матрица шага анализа, y_0^k – вектор данных наблюдений, x_k^f , x_k^a – значения, полученные на шаге прогноза и анализа, соответственно, H – оператор наблюдений:

$$y_0^k = H(x_t^k) + \varepsilon_0^k,$$

где x_t^k – «истинное» значение оцениваемой величины, ε_0^k – случайная ошибка наблюдений с нулевым математическим ожиданием и ковариационной матрицей R .

Считается, что на шаге прогноза оцениваемая переменная не меняется. Оценка производится для средних в заданном временном интервале значений. Отметим, что измеряемой переменной является концентрация метана на заданных уровнях по высоте. Для использования данных из большого временного интервала в оператор наблюдений включена модель переноса-диффузии.

В алгоритме LETKF на шаге анализа выполняется поиск оценки только для среднего по ансамблю значения, после чего вычисляется ансамбль значений анализов. Ниже представлены формулы, описывающие этап анализа алгоритма LETKF, для заданного момента времени k :

$$x_k^a = x_k^f + Dx_k^f \tilde{P}^a (HDx_k^f)^T R_k^{-1} [y_0^k - h(x_k^f)],$$

$$\tilde{P}^a = [(N-1)I + (HDx_k^f)^T R_k^{-1} HDx_k^f]^{-1},$$

где I – единичная матрица, N – размерность ансамбля, Dx_k^f – ансамбль ошибок прогноза, нормированных на $\sqrt{N-1}$, H – линейаризованный оператор наблюдений [16].

В рассматриваемом алгоритме для вычисления оценки производится прогноз по модели переноса-диффузии N элементов ансамбля прогнозов по заданному возмущению приземной концентрации. По ансамблю прогнозов вычисляется соответствующий ансамбль ошибок прогноза и оценка искомой величины.

В проводимых численных экспериментах с использованием данного алгоритма необходимо вводить контрольные механизмы для обеспечения стабильности и эффективности работы. С этой целью на этапе расчетов проводилась оценка следа ковариационной матрицы, и при несоответствии с установленным порогом применялся механизм коррекции (inflation factor). Этот механизм представляет собой процесс, при котором ковариационная матрица ошибок умножается на определенный коэффициент (в данном случае 1.1) до тех пор, пока матрица не будет удовлетворять требуемым критериям [16].

Организация численных экспериментов с модельными данными

Для численных экспериментов с модельными данными использовалась глобальная модель MOZART-4 [11–14]. В качестве реального («истинного») состояния системы были использованы данные о расчетах концентраций газов за 2016 г. Численное моделирование проводилось для всего Земного шара с использованием сетки модели с разрешением 2,8 градуса (стандартная сетка T-42).

Численные эксперименты с модельными данными являются общепринятой практикой и применяются для отладки системы усвоения данных и оценки ее эффективности. При этом задается «истинное» значение оцениваемых и моделируемых величин. Модельные данные получаются путем добавления к «истинному» значению случайных возмущений с нормальным распределением (с нулевым математическим ожиданием и заданной дисперсией), которые моделируют ошибки наблюдений в реальных данных. Для вычисления прогноза значение концентрации в приземном слое увеличивалось равномерно во всей поверхности Земного шара. Важно отметить, что как ошибка наблюдений, так и введенная погрешность концентрации метана в модельных экспериментах были выбраны и настроены таким образом, чтобы соответствовать смыслу физических аналогов в экспериментах с реальными данными.

В статье представлен численный эксперимент с модельными данными для процедуры усвоения данных. Процедура усвоения данных основана на алгоритме LETKF и работает циклично. Цикл начинается с расчета по модели на 7 дней, далее находится оценка приземной концентрации метана для средних значений по ансамблю за недельный период (7 дней).

Эта оценка усредняется для подобластей. Важно отметить, что алгоритм усвоения обладает свойством локальности, что позволяет ускорить вычисления и проводить их параллельно для каждой точки сетки. Усреднение для подобластей помогает сгладить случайные ошибки в наблюдениях, а также выявить динамику активности подобластей за длительный период, облегчая анализ больших объемов данных. Перед следующим циклом моделирования значения приземной концентрации метана обновляются, используя полученную оценку. Процесс моделирования продолжался циклично, с обновлением данных раз в неделю на протяжении всего 2016 г. Это позволило учесть долгосрочные тенденции в выделении метана и его влияние на распределении роста концентрации по всему Земному шару. Для оценки достоверности полученных результатов вычислялась норма отклонения оценки от принятого за реальное («истинное») состояния системы значения.

Также стоит подчеркнуть следующее.

- Выбор временного интервала в одну неделю для оценки приземной концентрации метана обусловлен необходимостью получения усредненных данных на достаточно короткий промежуток времени. Этот подход позволяет более точно оценивать динамику концентрации метана на поверхности Земли.
- Для обеспечения точности и надежности оценок приземной концентрации метана поверхность Земли была разбита на подобласти. Это позволило учесть различия в динамике и характеристиках регионов, а также сгладить возможные случайные ошибки и сбои в данных.
- Учитывая, что система усвоения данных настроена на оценку концентрации метана на нижней границе атмосферы, начальные значения концентрации метана для каждой подобласти были включены в анализ. Это обеспечило учет климатических особенностей и предварительную оценку концентрации метана в рассматриваемых регионах.
- Шаг анализа включал проведение экспериментов как для наблюдений, заданных в сетке модели, так и в сетке спутниковых наблюдений. Модельные данные наблюдений были интерполированы в узлы сетки, соответствующие спутниковым наблюдениям, для оценки влияния интерполяции на точность результатов. Это позволило оценить эффективность использования различных методов анализа и их влияние на конечные результаты.
- Применение локального алгоритма оценки позволило получать независимые оценки концентрации метана для каждого узла сетки модели.

Результаты численных экспериментов с модельными данными

С разработанным алгоритмом усвоения данных были проведены численные эксперименты с модельными данными. В первой серии экспериментов считалось, что модельные данные есть во всех узлах сетки модели на первом уровне по высоте каждые 6 часов.

На рис. 1 представлены результаты численных экспериментов с модельными данными. Горизонтальная ось отображает циклы усвоения данных, где каждый цикл соответствует одной неделе в течение года, что составляет 48 циклов усвоения за 2016 год. По вертикальной оси отмечается размер среднеквадратической ошибки получаемой оценки. Под среднеквадратической ошибкой понимается норма отклонения приземной концентрации после процедуры усвоения от «истинного» значения. Оранжевый график изображает введенную погрешность. Синий график соответствует результатам эксперимента, в котором происходит усвоение данных только на нижнем уровне модели, что соответствует симуляции усвоению данных наблюдений, доступных из сети наземных измерений. Оранжевый график представляет результаты эксперимента, в котором усваиваются данные наблюдений, заданных для вертикального профиля, аналогично данным, получаемым со спутника. Следует отметить, что второй вариант эксперимента имитирует усвоение данных наблюдений в каждой точке сетки и в каждый мо-

мент моделирования, однако в реальности такие наблюдения поступают в ограниченном объеме и количестве.

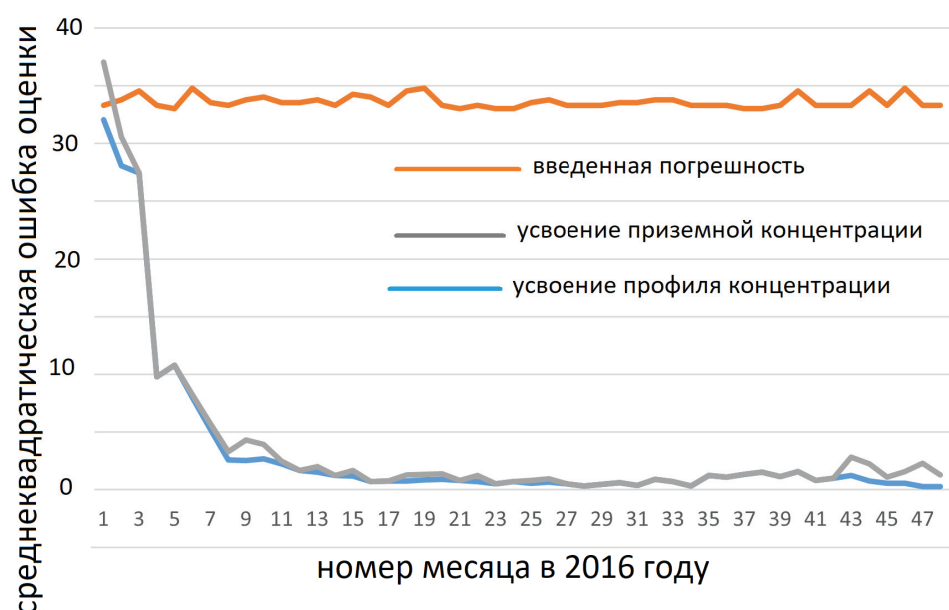


Рис. 1. Среднеквадратическая ошибка оценки приземной концентрации метана
Fig. 1. Root Mean Square Error in Estimating Surface Methane Concentrations

При введенной погрешности в размере 30–40 ppb получаемая среднеквадратическая ошибка оценки приземной концентрации метана меньше вносимой погрешности, что является достаточно хорошим показателем проведенного эксперимента. Отметим также, что в экспериментах с реальными данными достичь таких результатов невозможно, так как присутствует этап интерполяции данных, данные наблюдений доступны не в каждый момент времени и не в каждой точке сетки модели и т. д.

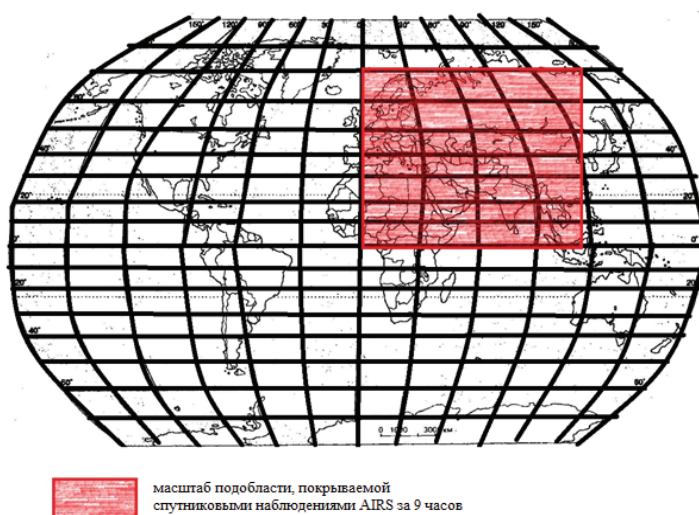


Рис. 2. Масштаб подобласти, покрываемой спутниковыми наблюдениями AIRS за окно усвоения (9 ч)

Fig. 2. Scale of subregion covered by AIRS satellite observations during the assimilation window (9 h)

Во второй серии экспериментов рассматривалось распределение данных в пространстве и времени, характерное для спутниковых данных AIRS [18; 19]. На рис. 2 представлен размер области, охватываемой спутниковыми измерениями, доступными для усвоения данных каждые 6 ч. В рамках алгоритма усвоения используется временной интервал (окно усвоения), в пределах которого используются данные наблюдений. В проведенных экспериментах окно усвоения составляет 9 ч, что означает, что данные наблюдений, полученные за 4,5 ч до и 4,5 ч после момента моделирования, учитывались при оценке. Рисунок иллюстрирует только масштаб (а не расположение) области, в которой можно получить оценки нижних границ концентрации метана с использованием спутниковых данных, каждые 6 ч. Каждые 6 ч подобласть, имеющая данные для усвоения, меняет свое положение, но размер остается примерно одинаковыми для каждого момента усвоения.

На рис. 3 представлены результаты численных экспериментов с использованием модельных данных, адаптированных под ошибки и ограничения, характерные для спутниковых данных AIRS. В случае сбоя измерительного инструмента или отсутствия данных AIRS на шаге анализа в определенной точке процесс поиска оценки приземной концентрации метана в этой точке не проводится. Вместо этого значение оценки, полученное на предыдущем временном шаге, остается неизменным и используется как текущая оценка для следующего шага по времени.

На рисунке приводится относительная ошибка (в процентах) оценок, найденных с 3-й по 48-ю неделю 2016 г. для каждой подобласти.

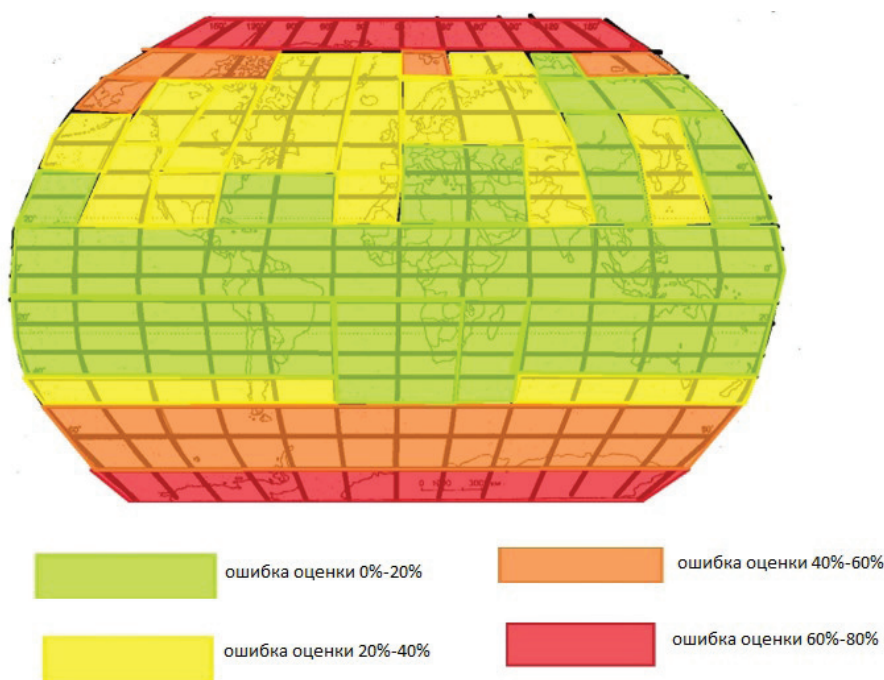


Рис. 3. Результаты модельных экспериментов с использованием модельных данных, адаптированных под характеристики спутниковых данных AIRS: средний процент ошибки оценки, получаемой с 3-й по 48-ю неделю 2016 г. по подобластям

Fig. 3. Results of model experiments using model data adapted to the characteristics of AIRS satellite data: average percentage of estimation error obtained from weeks 3 to 48 of 2016 by subarea

Как видно из рисунка, точность восстановления приземной концентрации зависит от региона. Это связано с тем, что спутник с измерительными приборами ориентирован по движению вдоль экватора, чем дальше от него область измерения, тем менее плотными становятся дан-

ные, получаемые со спутника. Также имеет место нерегулярность их поступления – данные для каждой подобласти поступают с интервалом примерно в 12 ч. Кроме того, на получаемые значения оказывают влияние наличие облаков и атмосферных аэрозолей, что может приводить к искажению данных.

На рис. 4 представлено поведение среднеквадратической ошибки получаемой оценки в результатах модельных экспериментов с учетом различных факторов, влияющих на условия эксперимента, проводимого с реальными данными наблюдений AIRS. Это включает эпизодичность поступления данных наблюдений AIRS, использование окна усвоения, наличие ошибок и другие аспекты. Также для сравнения на графике изображена введенная погрешность значений приземной концентрации. На рисунке приведены средние по всем подобластям на поверхности Земли значения, а также отдельно для подобласти «Сибирь» (50 градусов – 60 градусов с. ш., 80 градусов – 90 градусов в. д.).

Следует отметить, что на графике среднеквадратической ошибки оценки можно наблюдать ступенчатый характер улучшения получаемых оценок. Это указывает на постепенное снижение ошибки в процессе анализа данных и свидетельствует о том, что процедура усвоения данных, использующая модельные данные с учетом всех факторов, позволяет достичь достаточно высокой точности оценки приземной концентрации метана в атмосфере.

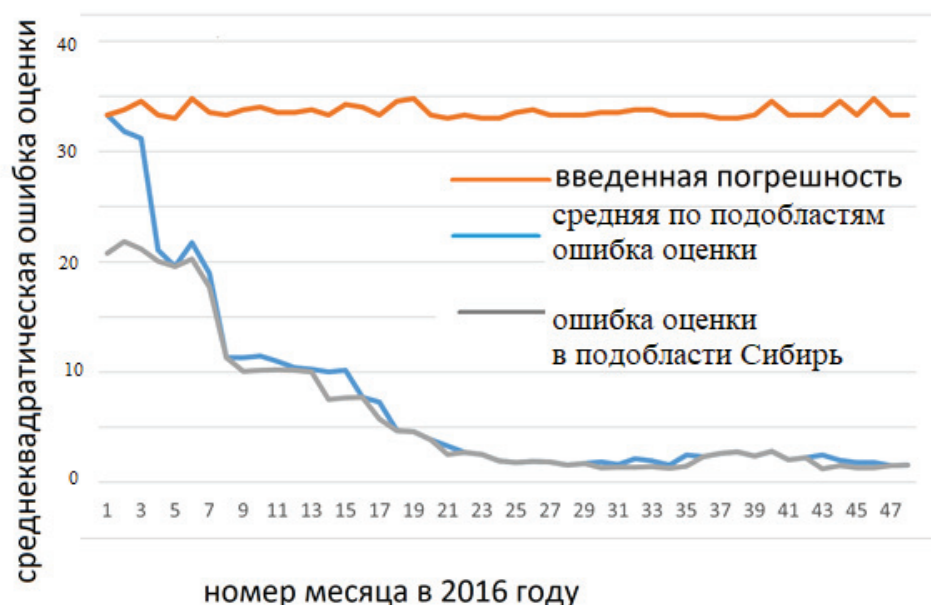


Рис. 4. График среднеквадратической ошибки оценки, получаемой при использовании имитации спутниковых данных AIRS

Fig. 4. Graph of the root mean square error of the estimate obtained using simulated AIRS satellite data

Таким образом, результаты численных экспериментов с модельными данными позволяют сделать вывод о возможности получения оценки приземной концентрации в процедуре усвоения данных с хорошей точностью. Отметим, что при проведении численных экспериментов с реальными данными полученная оценка приземной концентрации будет являться базой для оценки динамики потоков метана с поверхности Земли.

Заключение

В работе представлены результаты численных экспериментов с модельными данными по оценке приземной концентрации метана с использованием алгоритма ансамблевого фильтра Калмана. В алгоритме использовалась глобальная транспортная химическая модель MOZART-4. При моделировании данных наблюдений рассматривались различные сценарии их пространственно-временного распределения.

Из результатов проведенных экспериментов можно сделать вывод о том, что полученные оценки приземной концентрации метана достаточно точны и находятся в пределах заданной погрешности. Эти результаты говорят о возможности использования данного подхода для дальнейшей оценки эмиссии метана. Важно отметить, что анализ и оценка приземной концентрации метана являются ключевыми этапами для понимания процессов эмиссии парникового газа и разработки соответствующих стратегий уменьшения его выбросов.

Список литературы

1. Nakamura G., Potthast R. Inverse Modeling. IOP Publishing Ltd, 2015. P. 484.
2. Penenko A., Penenko V., Tsvetova E., Gochakov A., Pyanova E., Konopleva V. Sensitivity Operator Framework for Analyzing Heterogeneous Air Quality Monitoring Systems // Atmosphere. 2021. Vol. 12. P. 1–30.
3. Bocquet M. Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models // Atmos. Chem. Phys. Discuss. 2014. Vol. 14. P. 322.
4. Климова Е. Г. Экономичный алгоритм стохастического ансамблевого сглаживания // Сибирский журнал вычислительной математики. 2020. Т. 23, № 4. С. 381–394.
5. Климова Е. Г. Стохастический ансамблевый фильтр Калмана с трансформацией ансамбля возмущений // Сибирский журнал вычислительной математики РАН. 2019. Т. 22, № 1. С. 27–40.
6. Klimova E. G. Methods of estimation of greenhouse gases concentration in the atmosphere using observations and transport and diffusion model, based on the ensemble Kalman filter // All-Russian Conference “Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes, SDM”. 2017. Vol. 2033. P. 191–195.
7. Klimova E. G. Bayesian approach to data assimilation based on ensembles of forecasts and observations // IOP Conf. Series: Earth and Environmental Science, 2019. DOI 10.1088/1755-1315/386/1/012038
8. Feng L., Palmer P. I., Bosch H., Dance S. Estimating surface CO₂ fluxes from space-borne CO₂ dry air mole fraction observations using an ensemble Kalman filter // Atmospheric chemistry and physics. 2009. Vol. 9. P. 2619–2633.
9. Feng L. Evaluating a 3-D transport model of atmospheric CO₂ using ground-based, aircraft, and space-borne data // Atmospheric chemistry and physics. 2011. Vol. 11. P. 2789–2803.
10. Feng L., Palmer P. I., Parker R. J., Deutscher N. M., Feist D. G., Kivi R., Morino I., Sussmann R. Estimates of European uptake of CO₂ inferred from GOSAT X CO₂ retrievals: sensitivity to measurement bias inside and outside Europe // Atmospheric chemistry and physics. 2016. Vol. 16. P. 1289–1302
11. Platonova M. V., Klimova E. G. Determining the likely localization of methane sources using forecast time series and satellite data // CEUR Workshop Proceedings. 2021. Vol. 3006. P. 323–329.

12. **Platonova M. V., Klimova E. G.** An algorithm for estimating greenhouse gas fluxes using satellite data for a global transport and diffusion model // *Journal of Physics: Conference Series*. 2021. Vol. 1715(1). P. 012021.
13. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Methane emission by petroleum industry in Western Siberia according to satellite observations // *Atmospheric Physics*. 2021. Vol. 119166N. DOI 10.1117/12.2603438
14. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Regression model for reconstruction of the total methane content according to the data from AIRS hyperspectrometer and chemical transport model MOZART-4 // *Atmospheric Physics*. 2023. Vol. 1278062. DOI 10.1117/12.2690784
15. **Lu Xiao.** Methane emissions in the United States, Canada, and Mexico: evaluation of national methane emission inventories and 2010–2017 sectoral trends by inverse analysis of in situ (GLOBALVIEWplus CH₄ ObsPack) and satellite (GASAT) atmospheric observations // *Atmospheric Chemistry and Physics*. 2022. Vol. 22. P. 395–418.
16. **Bisht Jagat S. H.** Estimation of CH₄ emission based on an advanced 4D-LETKF assimilation system // *Geoscientific Model Development*. 2023. Vol. 16. P. 1823–1838.
17. **Evensen G.** Data assimilation. The ensemble Kalman filter. Berlin, Springer-Verlag. Heideberg, 2009. P. 307.
18. **Elansky N. F., Shilkin A. V., Ponomarev N. A., Zakharova P. V., Kachko M. D., Poliakov T. I.** Spatiotemporal Variations in the Content of Pollutants in the Moscow Air Basin and Their Emissions // *Izv. Atmos. Ocean. Phys.* 2022. Vol. 58, no. 1. P. 80–94.
19. **Kulawik S. S.** Evaluation of single-footprint AIRS CH₄ profile retrieval uncertainties using aircraft profile measurements // *Atmospheric Measurement Techniques*. 2021. Vol. 14. P. 335–354.

References

1. **Nakamura G., Potthast R.** Inverse Modeling, IOP Publishing Ltd, 2015, p. 484.
2. **Penenko A., Penenko V., Tsvetova E., Gochakov A., Pyanova E., Konopleva V.** Sensitivity Operator Framework for Analyzing Heterogeneous Air Quality Monitoring Systems. *Atmosphere*, 2021, vol. 12, pp. 1–30.
3. **Bocquet M.** Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmos. Chem. Phys. Discuss.*, 2014, vol. 14, p. 322.
4. **Klimova E. G.** Economical algorithm for stochastic ensemble smoothing. *Siberian Journal of Computational Mathematics*, 2020, vol. 23, no. 4, pp. 381–394. (in Russ.)
5. **Klimova E. G.** Stochastic ensemble Kalman filter with transformation of the ensemble of disturbances. *Siberian Journal of Computational Mathematics RAS*, 2019, vol. 22, no. 1, pp. 27–40. (in Russ.)
6. **Klimova E. G.** Methods of estimation of greenhouse gases concentration in the atmosphere using observations and transport and diffusion model, based on the ensemble Kalman filter. *All-Russian Conference "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes, SDM"*, 2017, vol. 2033, p. 191–195.
7. **Klimova E. G.** Bayesian approach to data assimilation based on ensembles of forecasts and observations. *IOP Conf. Series: Earth and Environmental Science*, 2019. DOI 10.1088/1755-1315/386/1/012038
8. **Feng L., Palmer P.I., Bosch H., Dance S.** Estimating surface CO₂ fluxes from space-borne CO₂ dry air mole fraction observations using an ensemble Kalman filter. *Atmospheric chemistry and physics*, 2009, vol. 9, pp. 2619–2633.
9. **Feng L.** Evaluating a 3-D transport model of atmospheric CO₂ using ground-based, aircraft, and space-borne data. *Atmospheric chemistry and physics*, 2011, vol. 11, pp. 2789–2803.

10. **Feng L., Palmer P. I., Parker R. J., Deutscher N. M., Feist D. G., Kivi R., Morino I., Sussmann R.** Estimates of European uptake of CO₂ inferred from GOSAT X CO₂ retrievals: sensitivity to measurement bias inside and outside Europe. *Atmospheric chemistry and physics*, 2016, vol. 16, pp. 1289–1302
11. **Platonova M. V., Klimova E. G.** Determining the likely localization of methane sources using forecast time series and satellite data. *CEUR Workshop Proceedings*, 2021, vol. 3006, pp. 323–329.
12. **Platonova M. V., Klimova E. G.** An algorithm for estimating greenhouse gas fluxes using satellite data for a global transport and diffusion model. *Journal of Physics: Conference Series*, 2021, vol. 1715(1), p. 012021.
13. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Methane emission by petroleum industry in Western Siberia according to satellite observations. *Atmospheric Physics*, 2021, vol. 119166N. DOI 10.1117/12.2603438
14. **Lagutin A. A., Mordvin E. Yu., Volkov N. V.** Regression model for reconstruction of the total methane content according to the data from AIRS hyperspectrometer and chemical transport model MOZART-4. *Atmospheric Physics*, 2023, vol. 1278062. DOI 10.1117/12.2690784
15. **Lu Xiao.** Methane emissions in the United States, Canada, and Mexico: evaluation of national methane emission inventories and 2010–2017 sectoral trends by inverse analysis of in situ (GLOBALVIEWplus CH₄ ObsPack) and satellite (GASAT) atmospheric observations. *Atmospheric Chemistry and Physics*, 2022, vol. 22, pp. 395–418.
16. **Bisht Jagat S. H.** Estimation of CH₄ emission based on an advanced 4D-LETKF assimilation system. *Geoscientific Model Development*, 2023, vol. 16, pp. 1823–1838.
17. **Evensen G.** Data assimilation. The ensemble Kalman filter. Berlin, Springer-Verlag. Heidelberg, 2009. P. 307.
18. **Elansky N. F., Shilkin A. V., Ponomarev N. A., Zakharova P. V., Kachko M. D., Poliakov T. I.** Spatiotemporal Variations in the Content of Pollutants in the Moscow Air Basin and Their Emissions. *Izv. Atmos. Ocean. Phys.*, 2022, vol. 58, no 1, pp. 80–94.
19. **Kulawik S. S.** Evaluation of single-footprint AIRS CH₄ profile retrieval uncertainties using aircraft profile measurements. *Atmospheric Measurement Techniques*, 2021, vol. 14, pp. 335–354.

Сведения об авторах

Платонова Марина Владимировна, младший научный сотрудник
Researcher ID WoS: AFI-8086-2022

Котлер Василий Дмитриевич, младший научный сотрудник

Климова Екатерина Георгиевна, доктор физико-математических наук
Researcher ID WoS: S-5641-2016

Information about the Authors

Marina V. Platonova, Junior Researcher
Researcher ID WoS: AFI-8086-2022

Vasily D. Kotler, Junior Researcher

Ekaterina G. Klimova, Doctor of Physical and Mathematical Sciences
Researcher ID WoS: S-5641-2016

*Статья поступила в редакцию 08.03.2024;
одобрена после рецензирования 05.04.2024; принята к публикации 05.04.2024*

*The article was submitted 08.03.2024;
approved after reviewing 05.04.2024; accepted for publication 05.04.2024*

Правила оформления текста рукописи

Авторы представляют статьи на русском или английском языке объемом от 0,5 авторского листа (20 тыс. знаков) до 1 авторского листа (40 тыс. знаков), включая иллюстрации (1 иллюстрация форматом 190×270 мм = 1/6 авторского листа, или 6,7 тыс. знаков). Публикации, превышающие указанный объем, допускаются к рассмотрению только после индивидуального согласования с редакцией журнала.

Текст рукописи должен быть представлен в редколлегию в виде файла MS Word (.doc, .docx). Гарнитура Times New Roman, размер шрифта 11, межстрочный интервал 1, размеры полей – стандартные значения текстового редактора. Форматирование – выравнивание по ширине страницы, переносы слов включены, каждый новый абзац начинается с красной строки. Не допускается ручное форматирование абзацев (пробелами, лишними переводами строк, разрывами страниц).

Структура статьи

- Индекс УДК (универсальной десятичной классификации). Выравнивание по левому краю
- Название статьи. Выравнивание по центру, полужирный шрифт
- ФИО авторов (полностью). Выравнивание по центру, полужирный шрифт
- Места работы всех авторов. Выравнивание по центру, курсив
- Адреса электронной почты, ORCID авторов
- Аннотация статьи
- Ключевые слова, не более 10
- Благодарности, сведения о финансовой поддержке
- Название статьи **на английском языке**. Выравнивание по центру, полужирный шрифт
- ФИО авторов **на английском языке** (полностью). Выравнивание по центру, полужирный шрифт
- Места работы авторов **на английском языке**. Выравнивание по центру, курсив
- Аннотация статьи **на английском языке (Abstract)**, 200–250 слов
- Ключевые слова **на английском языке (Keywords)**, не более 10
- Благодарности, сведения о финансовой поддержке **на английском языке**, если есть соответствующий раздел на русском языке (**Acknowledgements**)
- Основной текст
- Список литературы / **References**
- Сведения об авторах

Требования к оформлению основного текста и иллюстративных материалов

Основной текст должен быть представлен в структурированном виде, рекомендуется использовать подзаголовки – например: Введение, Методика..., Выводы, Результаты, Заключение.

Подзаголовки отделяются и набираются полужирным шрифтом. В целях выделения частей текста и отдельных слов и словосочетаний допускается использование курсива или полужирного шрифта. Подчеркивание, разрядка, изменение основного кегля и выделение цветом не используются.

Иллюстрации к рукописи статьи должны быть приложены в виде отдельных файлов. При этом в тексте должно содержаться включенное изображение с указанием имени файла. Все иллюстрации, содержащие схемы, графики, алгоритмы и т. п., должны быть представлены в векторном виде (.ai, .eps, .cdr). Скриншоты и другие растровые изображения должны быть представлены в максимально высоком качестве, без каких-либо потерь и искажений (.jpg, .tif). Все иллюстрации должны иметь подрисуночную подпись – свое название. Надписи к таблицам и подписи к иллюстрациям приводятся **на двух языках (русском и английском)**.

Примеры:

Рис. 1. Диаграмма производительности...

Fig. 1. Performance diagram...

Таблица 1

Сравнение алгоритмов...

Table 1

Comparison of algorithms...

Нумерация последовательная и неразрывная от начала статьи. Не допускается использование других наименований, кроме «Рис.» / «Fig.», «Таблица» / «Table», и усложнение нумерации (например, «Рис. 3.2.»). Ссылка на иллюстрацию в тексте должна быть приведена в круглых скобках, например: (рис. 1), (табл. 1).

Формулы должны быть набраны с использованием редактора MathType либо встроенного редактора формул MS Word. Кегль основных символов – 11, греческие символы набираются прямым шрифтом, латинские – курсивом. Нумеруются только те формулы, на которые автор ссылается в тексте.

Abstract

Аннотация статьи на английском языке (Abstract) не должна быть дословным переводом русскоязычной аннотации. Раздел Abstract, как и основной текст, должен быть структурирован, в нем должно содержаться описание цели работы, методов исследования, научной значимости, выводов / результатов. Требуется качественный перевод на английский язык (при необходимости просим авторов обращаться к профессиональным переводчикам). **Объем Abstract 200–250 слов.**

Список литературы / References

Список литературы и список литературы на английском языке (References) размещаются в общем разделе. Рекомендуемое количество цитируемых в статье источников – не менее 10, в список желательно включать ссылки на актуальные работы по теме исследования, особенно в иностранных периодических изданиях.

В тексте статьи ссылки на литературу указываются цифрами в квадратных скобках, при необходимости указываются номера страниц, например: [2; 3. С. 15].

Список литературы нумеруется в порядке цитирования и оформляется в соответствии с ГОСТ Р 7.0.5-2008 на библиографическое описание (знаки тире в описании опускаются). Ссылки на неопубликованные работы, а также на Интернет-ресурсы (кроме электронных изданий, поддающихся библиографическому описанию) оформляются в виде сноски.

В Список литературы ссылки на источники следует включать на оригинальном языке опубликования. Каждый источник должен быть также оформлен на английском языке (References) по международному стандарту для публикаций в области информатики IEEE Style со следующими отличиями:

- инициалы авторов указываются после фамилии;
- название статьи не берется в кавычки, отделяется точкой;

- отсутствует союз «and» перед фамилией последнего автора;
- в диапазоне страниц – удвоенная «р» (например, «pp. 2–9»);
- год издания указывается после места издания (для книг) и сразу после названия журнала (для периодики).
- Перевод источника на английский язык:
- если источник имеет выходные данные на английском языке, то для формирования References **следует использовать именно эти данные**;
- если оригинальная публикация не содержит выходных данных на английском языке, то допускается транслитерация названия материала на латинский алфавит в сочетании с переводом на английский язык в квадратных скобках. В конце описания указывается, на каком языке написана эта работа, например, (in Russ.). При транслитерации можно воспользоваться интернет-ресурсом <http://ru.translit.ru/>, рекомендуется выбрать стандарт BSI. Место издания не транслитерируется, указывается полностью на английском языке, например: Moscow. Название издательства / издателя, как правило, транслитерируется. Для журналов, у которых есть официальное название на английском языке, – использовать его (проверить на сайте журнала, или, например, в библиотеке WorldCat), если названия на английском языке нет, использовать транслитерацию по системе BSI. Не следует самостоятельно переводить названия журналов.

Если у цитируемого источника есть **цифровой идентификатор DOI** (<https://search.crossref.org/>), его требуется обязательно указывать в конце библиографической ссылки.

Примеры оформления ссылок. Каждый источник в том же пункте дублируется на английском языке (References).

Источник на русском языке, перевод на английский доступен в метаданных статьи

1. Журавлев С. С., Рудометов С. В., Окольников В. В., Шакиров С. Р. Применение модельно-ориентированного проектирования к созданию АСУ ТП опасных промышленных объектов // Вестник НГУ. Серия: Информационные технологии. 2018. Т. 16, № 4. С. 56–67. DOI 10.25205/1818-7900-2018-16-4-56-67

Zhuravlev S. S., Rudometov S. V., Okolnishnikov V. V., Shakirov S. R. Model-Based Design Approach for Development Process Control Systems of Hazardous Industrial Facilities. *Vestnik NSU. Series: Information Technologies*, 2018, vol. 16, no. 4, pp. 56–67. (in Russ.) DOI 10.25205/1818-7900-2018-16-4-56-67

Источник на английском языке. Оформляем согласно требованиям для References. Приводим только 1 раз.

2. Telnov V. I. Optimization of the Beam Crossing Angle at the ILC for E + e- and yy Collisions. *Journal of Instrumentation*, 2018, vol. 13, no. 03, pp. P03020–P03020. DOI 10.1088/1748-0221/13/03/p03020

Метаданные источника доступны только на русском языке

3. Жижимов О. Л., Федотов А. М., Шокин Ю. И. Технологическая платформа массовой интеграции гетерогенных данных // Вестник НГУ. Серия: Информационные технологии. 2013. Т. 11, вып. 1. С. 24–41.

Zhizhimov O. L., Fedotov A. M., Shokin Yu. I. Tekhnologicheskaya platforma massovoi integratsii geterogennykh dannykh [Technology Platform for the Mass Integration of Heterogeneous Data]. *Vestnik NSU. Series: Information Technologies*, 2013, vol. 11, no. 1, pp. 24–41. (in Russ.)

Сведения об авторах

Последний раздел статьи – информация об авторе / авторах **на русском и английском языках**:

- ФИО полностью, ученая степень, ученое звание;
- идентификаторы автора, такие как ResearcherID (всем авторам рекомендуется использовать данные сервисы для ведения актуального списка своих публикаций);
- контактный телефон (не публикуется).

Если статья представляется на английском языке, необходимо приложить перевод на русский язык названия, аннотации, ключевых слов, сведений об авторе.

Доставка материалов

Материалы предоставляются в редакцию по электронной почте inftech@vestnik.nsu.ru.

Порядок рецензирования

Все статьи сначала проходят проверку на заимствование и только после этого отправляются на рецензирование. Редакционный совет не допускает к публикации материал, если имеется достаточно оснований полагать, что он является плагиатом.

Тип рецензирования статей – двухуровневое, одностороннее анонимное («слепое»).

Для каждой статьи редколлегией выбираются рецензенты, научная деятельность которых связана с темой представленного материала. Ответственный секретарь журнала обращается к ним с просьбой дать экспертную оценку статье либо помочь организовать рецензирование.

Рецензии для журнала «Вестник НГУ. Серия: Информационные технологии» составляются по единой схеме и подразумевают оценку по следующим критериям: соответствие тематике журнала, оригинальность и значимость результатов, качество изложения материала.

Заполненный бланк рецензии высылается на электронный адрес редакции. В зависимости от экспертных заключений статья может быть принята редакционным советом к опубликованию, рекомендована автору к доработке (с последующим повторным рецензированием либо без него) или отклонена (с предоставлением автору мотивированного отказа). Автору на электронный адрес высылается текст рецензии без указания ФИО рецензента и его контактных данных.

Все рецензии хранятся в редакции журнала не менее 5 лет. Редколлегия журнала обязуется при поступлении соответствующего запроса направлять копии рецензий в Министерство науки и высшего образования Российской Федерации.

Вестник Новосибирского государственного университета
Серия: Информационные технологии
Научный журнал
2024. Т. 22, № 1

Учредитель
Новосибирский государственный университет

Адрес учредителя, издателя
ул. Пирогова, 2, Новосибирск, 630090, Россия

ВрИО главного редактора
д-р физ.-мат. наук, проф. М. М. Лаврентьев

Адрес редакции
ул. Пирогова, 1, Новосибирск, 630090, Россия

Дата выхода в свет 05.07.2024
Тираж 27 экз.
Цена 659 руб.

Адрес типографии
ул. Пирогова, 2, Новосибирск, 630090, Россия

Подписной индекс
Пресса России – 11226

ISSN 1818-7900



2 2 0 0 1

