ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ, НГУ)
Факультет информационных технологий
Кафедра параллельных вычислений
09.03.01 Информатика и вычислительная техника. Программная инженерия и компьютерные науки

Разработка и реализация программных средств автоматизированного обнаружения семантических ошибок во фрагментированных программах для системы LuNA

Выполнил: Мичуров Михаил Антонович, студент группы 18201 ФИТ НГУ

Научный руководитель: Власенко Андрей Юрьевич, к.т.н., доц. каф. ПВ ФИТ НГУ

Рецензент: Власов Александр Александрович, к.т.н., с.н.с. ИНГГ СО РАН

Новосибирск 2022

20.06.2022

Актуальность

Система конструирования параллельных программ численного моделирования **LuNA,** разрабатываемая в ИВМиМГ СО РАН, представляется перспективной и актуальной для высокопроизводительных вычислений.

Из-за специфики парадигмы фрагментированного программирования, реализуемой системой LuNA и специфических семантических ошибок во фрагментированных программах и**спользование традиционных средств отладки затруднено**.

В связи с этим **актуальной является разработка** специализированного **средства автоматизированной отладки** LuNA-программ.

Цель и задачи

ЦЕЛЬ:

Разработка модуля сбора трассы для системы LuNA и программного средства анализа трассы с целью обнаружения семантических ошибок.

ЗАДАЧИ:

1. Провести обзор и анализ существующих методов отладки параллельных программ, а также программных средств, реализующих данные методы.
2. Провести анализ предметной области и выделить классы семантических ошибок, свойственных фрагментированным программам.
3. Сформулировать требования к разрабатываемому программному средству.
4. Разработать и реализовать обнаружение наиболее типичных для фрагментированных программ семантических ошибок в программном средстве.
5. Протестировать разработанное средство на ряде фрагментированных программ и оценить накладные расходы.

Система LuNA

LuNA (Language for Numerical Algorithms) – система фрагментированного программирования, разрабатываемая в ИВМиМГ СО РАН;

LuNA-программа - описание фрагментированного алгоритма на языке LuNA как множества **фрагментов данных** и **фрагментов вычислений**;

Фрагменты вычислений (ФВ) - независимые единицы программы, которые содержат описание входных/выходных фрагментов данных и кода фрагмента;

Фрагмент данных (ФД) – агрегат из переменных.

Семантические ошибки во фрагментированных программах

- Использование неинициализированных данных
- Гонки данных
- Несоответствие типов

Существующие методы и средства отладки

- Диалоговая отладка (TotalView, Distributed Debugging Tool)
- Верификация модели программы (MPI-Spin)
- Сравнительная отладка (DVM)
- Автоматизированный анализ корректности
  - Анализ во время работы программы (Intel Trace Analyzer
  - and Collector)
  - **Анализ по трассе** (Intel Message Checker)
- ...

Используемые методы и технологии

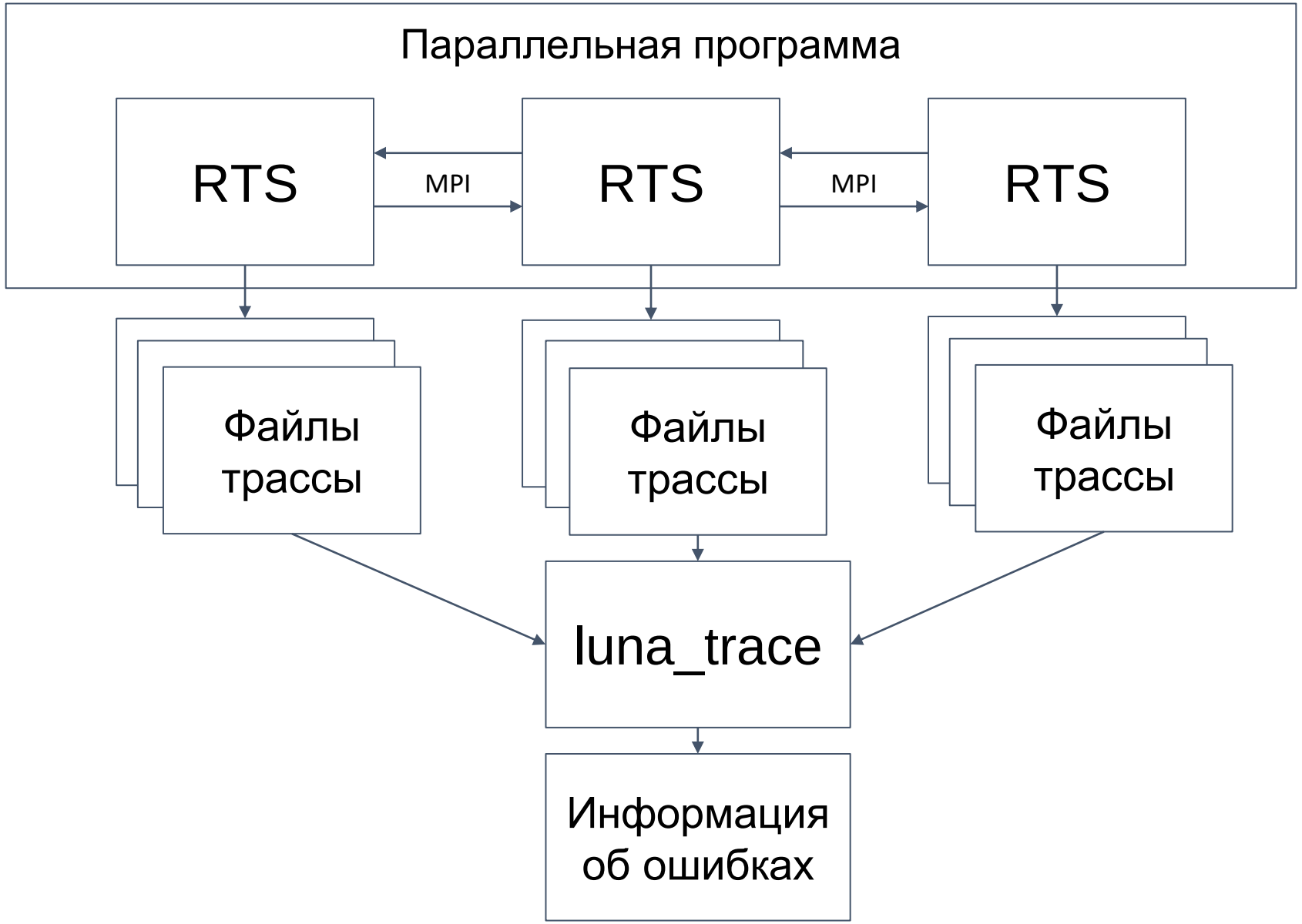
Разработанное программное средство реализует подход “анализ по трассе” (“посмертный анализ”)

- Требует минимального вмешательства в систему LuNA
- Предоставляет простор в выборе языков и технологий, используемых при реализации

Для разработки средства отладки выбран язык Python

- Используется в некоторых компонентах системы LuNA
- Компромисс между скоростью разработки и скоростью выполнения

Схема работы созданных средств



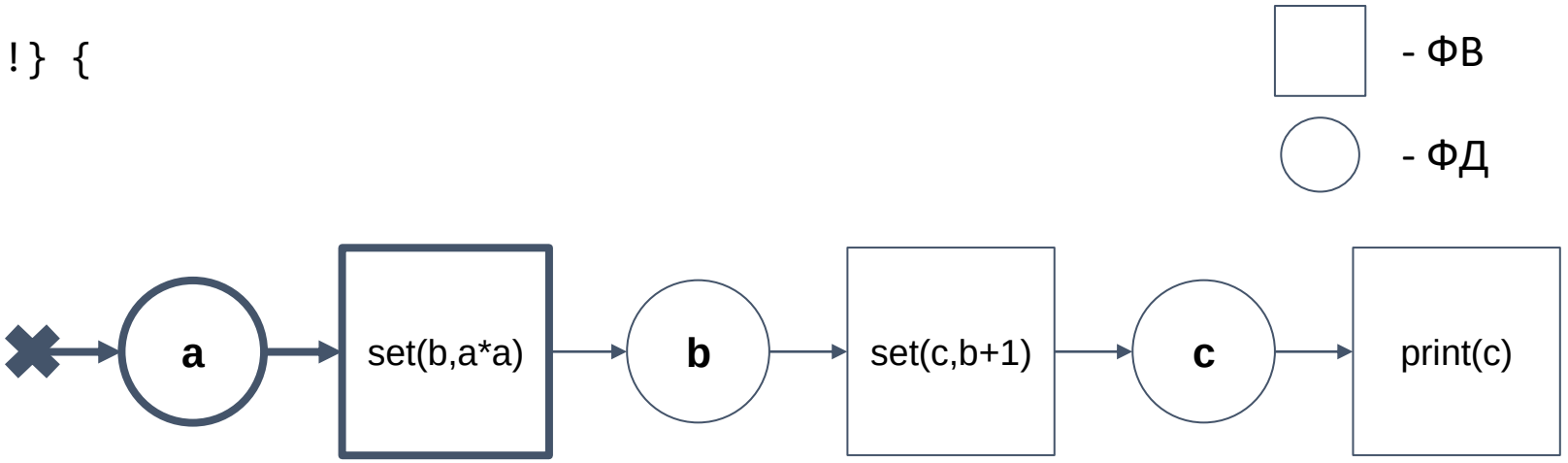
Пример некорректной LuNA-программы (использование неинициализированных данных)

root.fa

```
C++ sub set(name x, int v) ${<!} {
  x.setValue<int>(v);
}!

C++ sub print(int v) ${<!} {
  printf("%d\n", v);
}!

sub main() {
  df a, b, c;
  set(b, a * a);  // Данный ФВ не может быть выполнен и приведет к зависанию
  set(c, b + 1);
  print(c);
}
```



Информация о причинах зависания

luna_trace предоставляет следующую информацию:

- выражение, вызвавшее зависание;
- номер строки и имя файла для выражения, вызвавшего зависание;
- стек вызовов процедур до этого выражения;
- информация о данных, отсутствие которых вызвало зависание;
- информация о местах в коде программы, где могла быть пропущена инициализация.

```
Following CFs appear to be the root cause of the system hang:

set(b, a * a) [./root.fa:11] never finished
main
    in sub main() [./root.fa:9]

    Awaited DF a
        full name a, a declared in sub main
 -  -  -  -

2 more CFs never finished but had data dependencies on other hang CFs
```

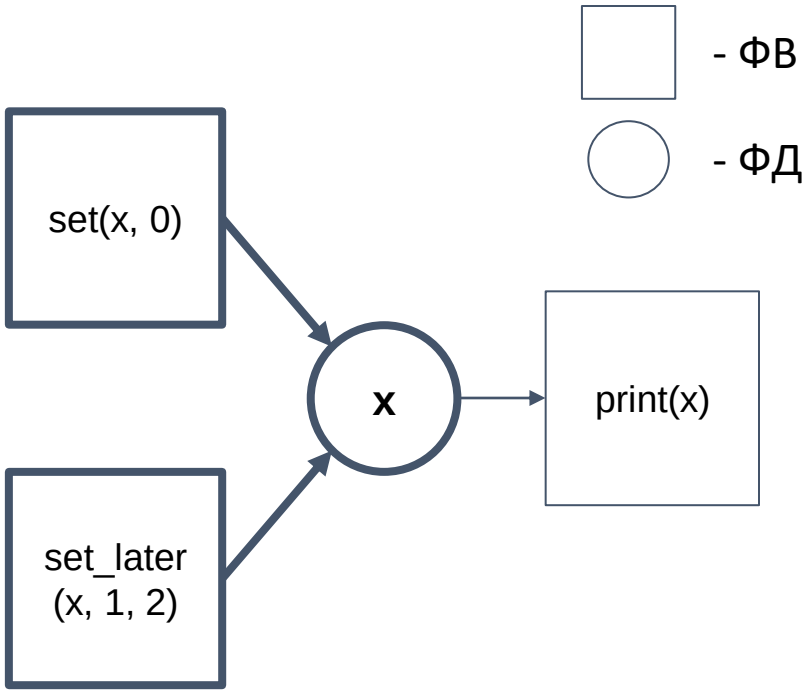
Пример вывода luna_trace для предыдущей программы

Пример некорректной LuNA-программы (повторная инициализация/гонка данных)

main.fa

```
import c_init_later(name, int, int) as set_later;
C++ sub set(name x, int val) ${<!} {
    x.setValue<int>(val);
    printf("Initialized DF to %d\n", x.getValue<int>());
}!
C++ sub print(int x) ${<!} { printf("%d\n", x); }!
sub call_set_later(name v) { set_later(v, 1, 2); }

sub main() {
    df x;
    set(x, 0);          // Инициализация x
    call_set_later(x); // Инициализация x
    print(x) @ {
        delete x;
    };
}
```



Информация о повторной инициализации

luna_trace предоставляет следующую информацию:

- имена ФД, инициализация которых происходит несколько раз;
- выражения, выполняющие повторную инициализацию;
- номер строки и имя файла для выражений, выполняющих повторную инициализацию;
- стек вызовов процедур до этого выражения.

```
Following DFs are initialized multiple times:

x (declared in sub main) in
    set(x, 0) [./main.fa:30]
        in sub main() [./main.fa:27]

    set_later(v, 1, 2) [./main.fa:13]
        in sub call_set_later(name v) [./main.fa:12]
        in call_set_later(x) [./main.fa:31]
        in sub main() [./main.fa:27]


 -  -  -  -
```

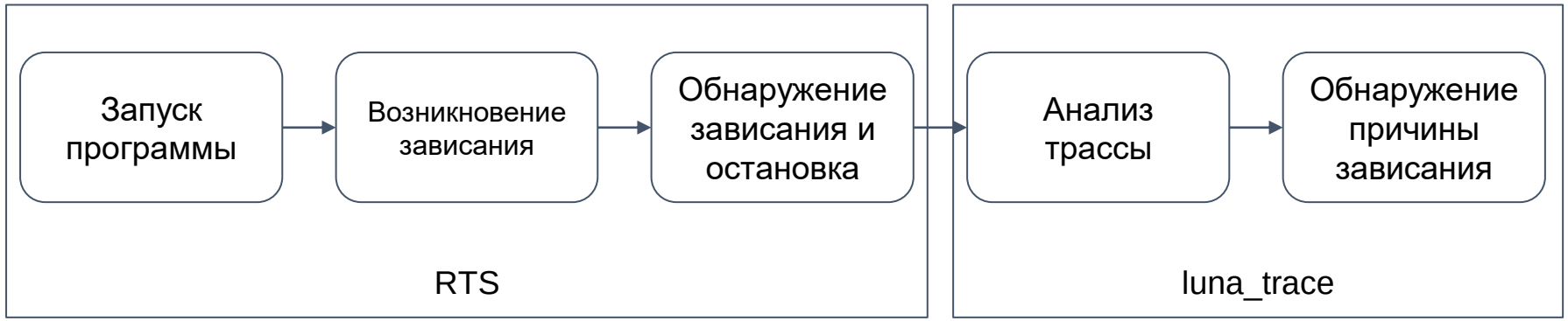
Пример вывода luna_trace для предыдущей программы

Остановка в ситуации зависания

На основе существующего модуля остановки системы (реализующего адаптированную версию алгоритма Дейкстры-Шольтена) реализована возможность остановки при зависании, вызванном использованием неинициализированных данных.

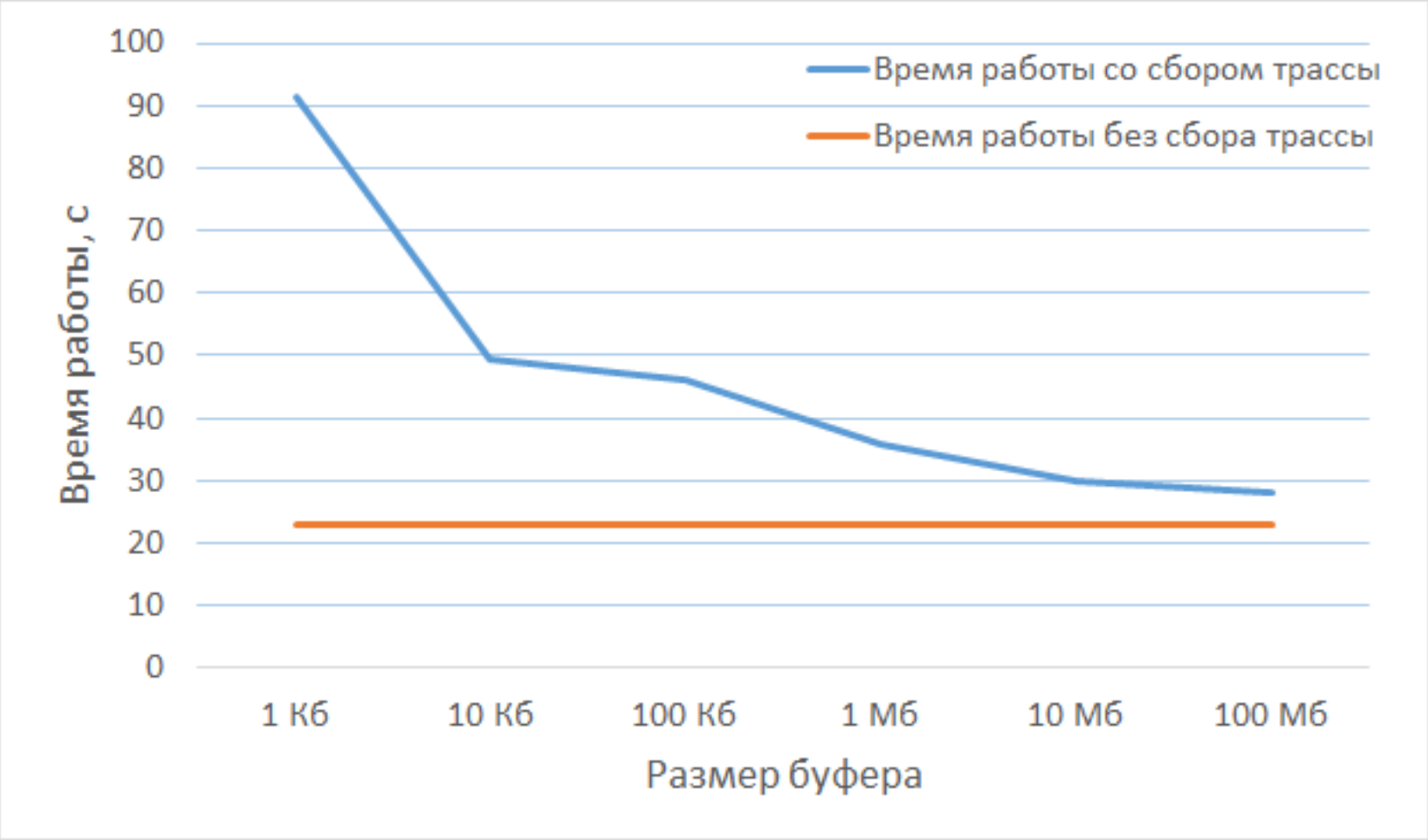
При обнаружении зависания пользователю **выводится соответствующее сообщение**, после чего можно использовать разработанное средство отладки для поиска причин зависания и их исправления

```
err> 0 ERROR:  System is idle, but unfinished CFs (3) exist. ./src/rts/rts.cpp:140
```



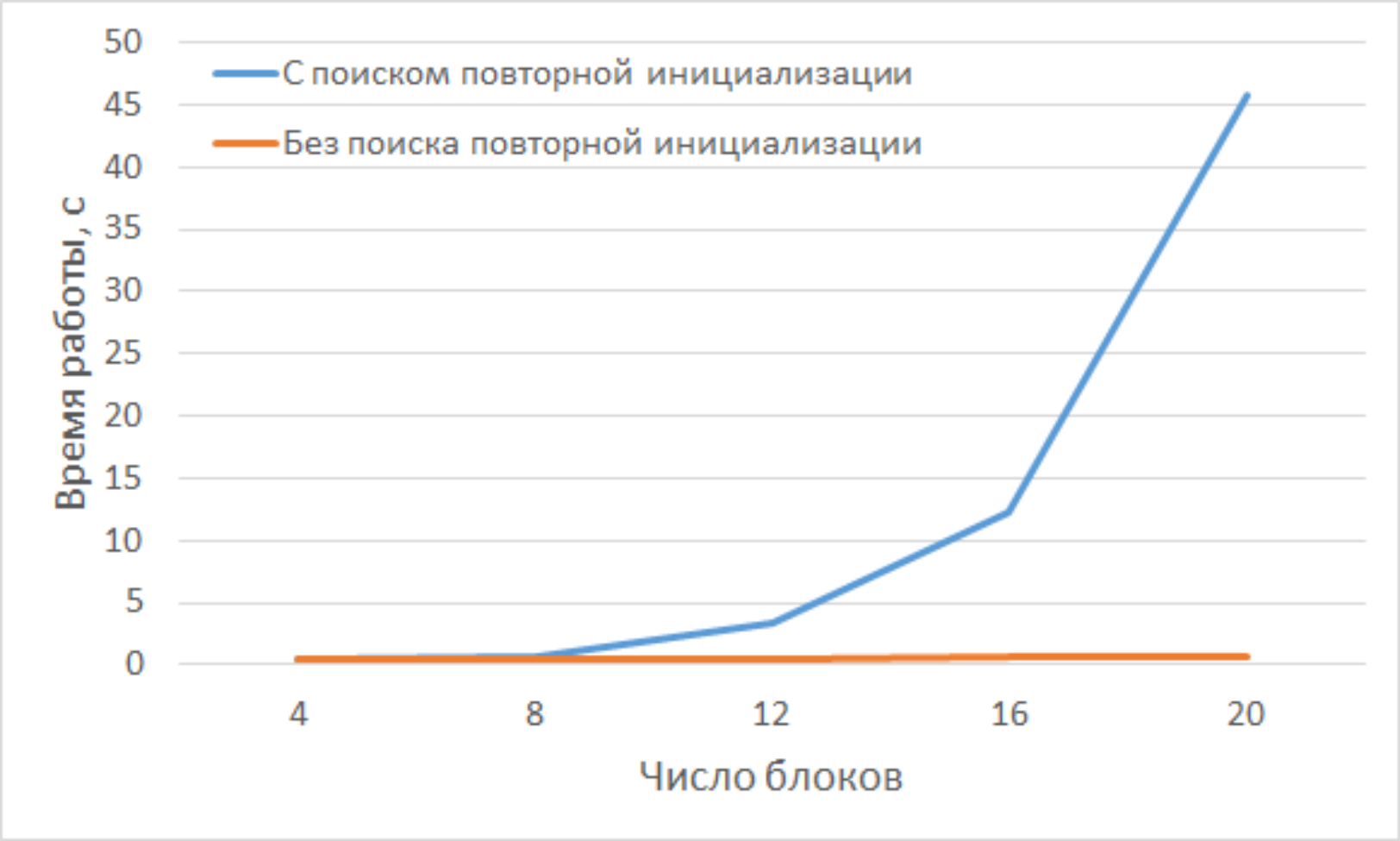
Оценка накладных расходов при сборе трассы

Время работы программы умножения квадратных матриц 2000x2000 (размер блока 50x50) в зависимости от размера буфера файлов трассы



Время анализа трассы

Время работы luna_trace в зависимости от числа блоков (блочное умножение матриц размером 2000 на 2000 с варьируемым размером блока)



Результаты

Разработан **модуль трассировки** для системы LuNA с вариабельным размером буфера трассы.

- В систему LuNA добавлена возможность **остановки при обнаружении зависания**, вызванного ошибкой, связанной с использованием неинициализированных данных.

Разработанная **утилита анализа трассы** способна предоставлять информацию о:

- причинах зависания системы;
- повторной инициализации данных.

Разработанное средство отладки было протестировано на нескольких некорректных LuNA-программах.

Проведено тестирование с целью оценки накладных расходов.

Заключение

Научная новизна состоит в анализе семантических ошибок, свойственных фрагментированным программам, а также в реализации метода автоматизированного контроля корректности по собранной трассе применительно к фрагментированным программам.

Практическая ценность работы обусловлена тем, что использование разработанного средства может существенно упростить разработку и отладку программ для системы LuNA.

Перспективность:

Разработанное средство в будущем может быть расширено функциональностью по обнаружению дополнительных классов семантических ошибок, а также другими методами и подходами к отладке.

Публикации

1. Мичуров М. А. Средство анализа причин зависаний фрагментированных программ в системе LuNA // Инновации. Наука. Образование. 2021. № 40. С. 354-364.
2. Власенко А. Ю., Мустафин Д. Э., Мичуров М. А. Автоматизация отладки и оптимизации фрагментированных программ в системе LuNA // (статья принята к опубликованию и будет напечатана в №3 журнала “Проблемы информатики” за 2022 год)
3. Мичуров М. А. Автоматизация обнаружения семантических ошибок во фрагментированных программах для системы LuNA // Информационные технологии: Материалы 60-й Междунар. науч. студ. конф. 10–20 апреля 2022 г.

Апробация

1. МНСК-2022 // Секция “Параллельные вычисления”
2. Всероссийская летняя XXXVII молодежная Школа-конференция по параллельному программированию

Спасибо за внимание!

ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ГЕОФИЗИКИ СО РАН

ИВМ и МГ СО РАН
пр. Лаврентьева 6
630090, Новосибирск,
РОССИЯ

ПРОБЛЕМЫ ИНФОРМАТИКИ

Тел: +7-383-330-96-43
Mail to: problem-info@sscc.ru
http: problem-info.sscc.ru

СПРАВКА

Настоящим уведомляю, что статья А.Ю. Власенко, М.А. Мичурова, Д.Э. Мустафина «Автоматизация отладки и балансировки нагрузки во фрагментированных программах» была принята к опубликованию и будет напечатана в № 3 журнала «Проблемы информатики» за 2022 год.

Главный редактор
журнала «Проблемы информатики»,
Д. т. н., профессор

В. Э. Малышкин

Для документов
(ИВМиМГ СО РАН)

Семантическая ошибка

Появление одного или более некорректных результатов во время работы программы.

Корректность программы

Отсутствие ошибочного поведения при работе программы.