МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ, НГУ)

Факультет информационных технологий
Кафедра параллельных вычислений

Направление подготовки 09.03.01 Информатика и вычислительная техника
Направленность (профиль): Программная инженерия и компьютерные науки

РАЗРАБОТКА И РЕАЛИЗАЦИЯ ПОДСИСТЕМЫ ПРОФИЛИРОВАНИЯ ДЛЯ СИСТЕМЫ ФРАГМЕНТИРОВАННОГО ПРОГРАММИРОВАНИЯ LUNA

Выполнила: Абрамушкина Екатерина Сергеевна, студентка группы 19202 ФИТ НГУ
Научный руководитель: Маркова Валентина Петровна, д.т.н., доцент, доц. каф. ПВ ФИТ НГУ
Соруководитель: Киреев Сергей Евгеньевич, ст. преп. каф. ПВ ФИТ НГУ
Рецензент: Вяткин Юрий Викторович, руководитель отдела Биоинформационных технологий, ООО “Новые Программные Системы”

Новосибирск 2023

23.06.2023

Введение в предметную область

Профилирование – это сбор характеристик исполнения программы, которые впоследствии используются для исследования поведения программы.

Существуют различные подходы к профилированию:

- **статический** – на основе анализа исходного кода, не требуется запуск программы;
- **динамический** – данные анализируются в процессе исполнения программы;
- **посмертный (post mortem)** – характеристики вычисляются на основе анализа трассы.

Трассировка – это запись без промежуточной обработки информации о текущем состоянии программы во время ее исполнения.

Система программирования LuNA

LuNA – это система автоматического конструирования параллельных программ разрабатываемая в ИВМиМГ СО РАН.

- Task-based система с крупноблочным параллелизмом.
- Реализует технологию фрагментированного программирования.
  - **Фрагментированная программа** – это множества переменных (фрагментов данных) и фрагментов вычислений.
  - **Фрагменты вычислений** – это независимые единицы программы, которые содержат описание входных и выходных фрагментов данных и кода фрагмента.
  - **Фрагмент данных** – это неизменяемый блок данных заданного размера.

Актуальность проблемы

Задачи, для которых в настоящее время используется профилирование в системе программирования LuNA:

- Повышение производительности системы LuNA.
- Обнаружение ошибок в сконструированных программах.

В дальнейшем сферы применения профилирования в системе LuNA будут расширяться.

Существующие средства профилирования, которые будут рассмотрены далее в обзоре, не удовлетворяют потребностям и имеют ряд проблем, которые нуждаются в решении.

Цель работы и задачи

Цель работы: разработать и реализовать подсистему профилирования в системе фрагментированного программирования LuNA.

Задачи для достижения цели:

- изучить существующие в системе средства профилирования фрагментированных программ;
- составить и проанализировать список требований, которым должна удовлетворять подсистема профилирования;
- спроектировать архитектуру подсистемы профилирования на основе поставленных требований;
- реализовать и протестировать прототип подсистемы профилирования.

Обзор средств профилирования программ

Сторонние средства профилирования параллельных программ:

- MPE
- Extrae, Paraver
- Pablo Performance Analysis Toolkit
- Vampir, VampirTrace
- Intel Trace Analyzer and Collector
- Scalasca (KOJAK)
- TAU (Tuning and Analysis Utilities)
- HPCToolkit

Средства предоставляют пользователю низкоуровневую информацию о работе программы, которую трудно соотнести с фрагментами данных и фрагментами вычислений в фрагментированных программах.

Обзор средств профилирования программ

Средства профилирования в системе LuNA:

- макросы LOG, TRACE, …
- WorkflowTracer
- LuNA Evlog, LuNA Evlog Analyzer

Средства представляют разрозненный набор инструментов, каждый был разработан для конкретной задачи и имеет ограниченную область применения. Несовместимы между собой, частично дублируют друг друга, нет возможностей для расширения и переиспользования функциональности.

Требования к подсистеме профилирования

Функциональные требования:

- Сбор трассы программы.
- Сбор профиля программы.
- Накопление информации об исполняемой программе во время ее исполнения для использования другими подсистемами.
- Восстановление последовательности событий исполнения параллельной программы по трассе.

Нефункциональные требования:

- Расширяемость функциональности подсистемы.
- Минимизация накладных расходов на работу подсистемы.
- Удобный и понятный интерфейс.
- Наличие технической документации.

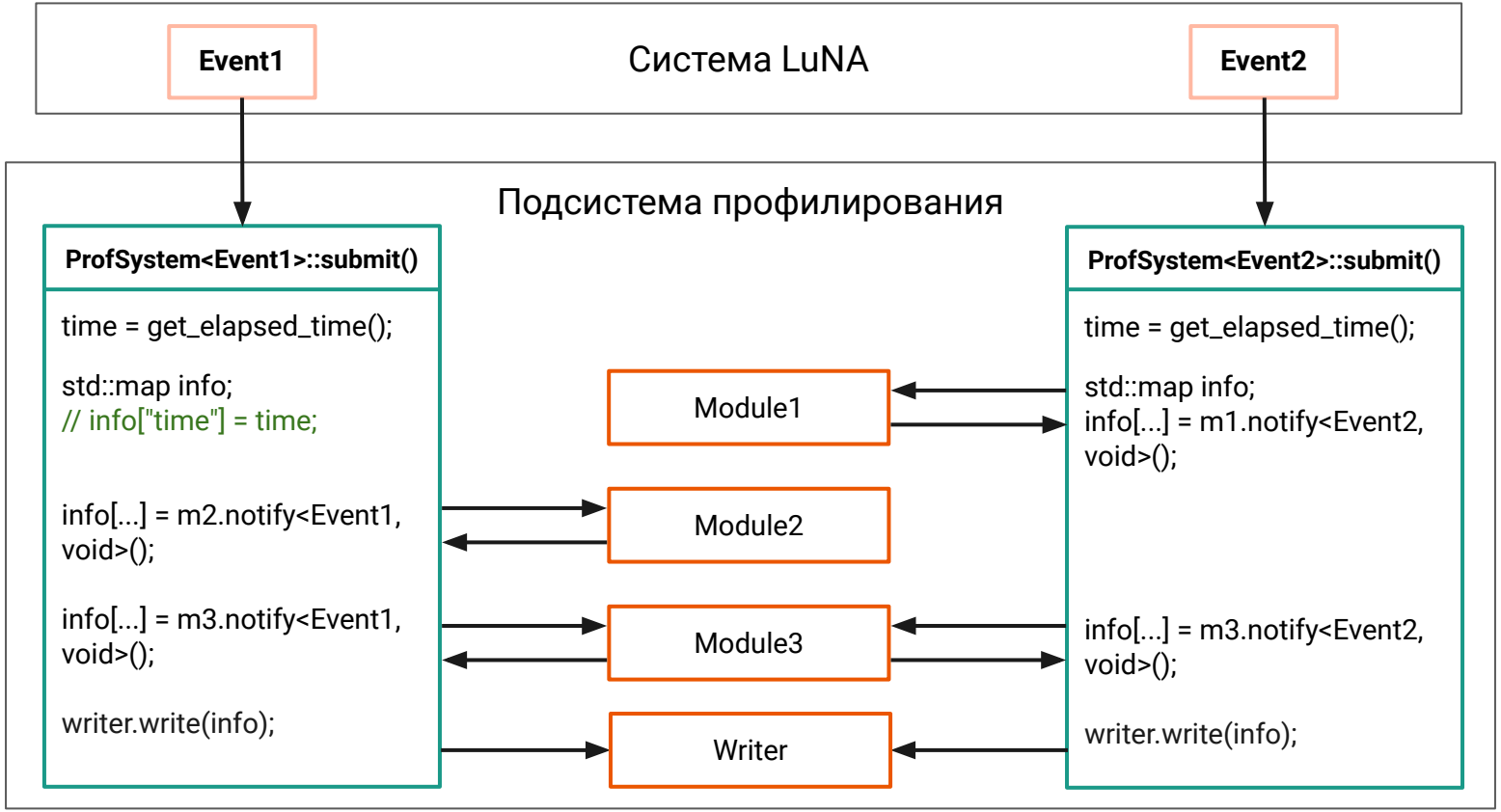
Предлагаемое решение

1. **Использовать событийно-ориентированную архитектуру.**
   События – это существенные изменения состояния системы. Подсистема должна реагировать на события, то есть существенные изменения состояния системы, и сохранять информацию о событиях в профиле исполнения программы.

2. **Использовать модульную архитектуру.**
   Подсистема профилирования – это набор независимых компонентов (модулей), которые взаимодействуют между собой.

3. **Предусмотреть методы для минимизации накладных расходов на работу подсистемы.**
   - Выбирать необходимые модули при компиляции системы.
   - Включать модули при запуске системы с помощью интерфейса командной строки.

Разработанная архитектура подсистемы



Реализация прототипа

- Прототип является MVP. Он встроен в существующий проект – модуль балансировки нагрузки в системе на основе трассировки.
- Для реализации программного интерфейса подсистемы использован класс ProfSystem. Данный класс написан с использованием шаблонов в языке C++ и имеет частичные специализации.
- Был также реализован модуль TraceWriter, который осуществляет запись трассы в файл.
- Был реализован алгоритм Кристиана для синхронизации часов.

Тестирование

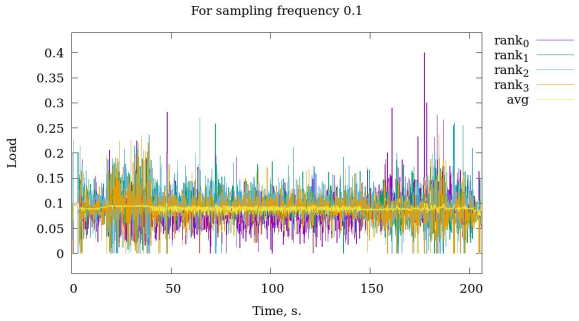
- Проводилось на ноутбуке Lenovo Legion 5P 15IMH05, процессор Intel Core i5-10300H 2.50GHz, 4 ядра, 8 Гб RAM.
- Программа, реализующая алгоритм блочного умножения квадратных матриц. Размер матрицы задается размером блока и количеством блоков.
- Тестирование проводилось для матриц размерностью 20×20 блоков, каждый блок является подматрицей размера 200×200. Тип данных – число с плавающей точкой.

Тестирование

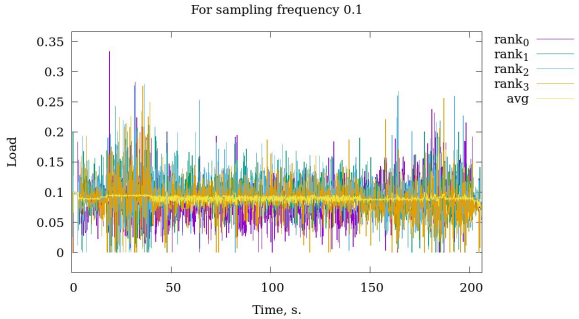
1. Тестирование работоспособности прототипа подсистемы

Цель теста: проверить, что прототип работоспособен и обеспечивает необходимую функциональность.

Трассировка при помощи макроса → Модуль балансировки нагрузки в системе LuNA →



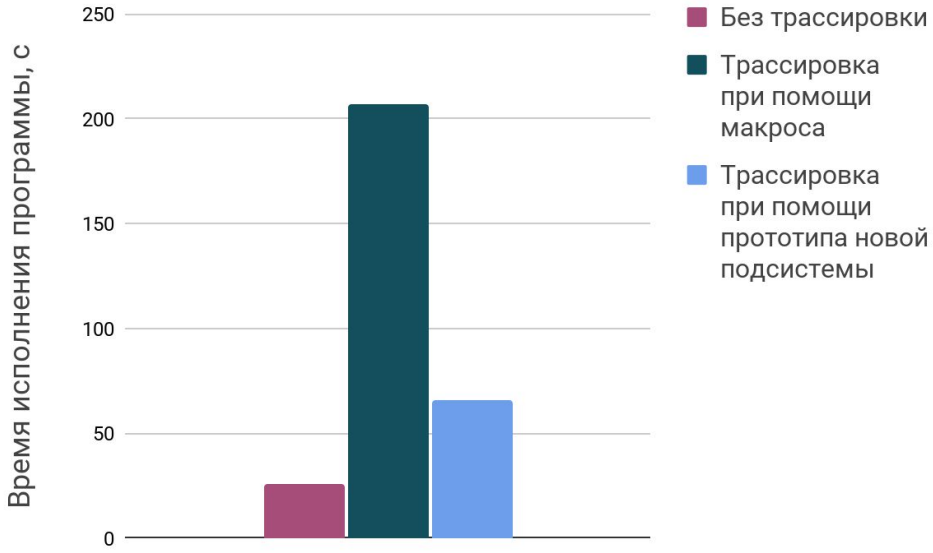
Трассировка при помощи прототипа → Модуль балансировки нагрузки в системе LuNA →



Тестирование

2. Исследование доли накладных расходов на работу прототипа подсистемы

Цель теста: сравнить накладные расходы на трассировку программы при помощи макроса и прототипа новой подсистемы.



Апробация

- Всероссийская летняя XXXIX молодежная Школа-конференция по параллельному программированию с международным участием (4 – 15 июля 2022 г.).

  Абрамушкина Е.С. Проектирование подсистемы профилирования для системы фрагментированного программирования LuNA (принято в печать).

- 61-й Международная научная студенческая конференция (17 – 26 апреля 2023 г.).

  Абрамушкина Е.С. Разработка подсистемы профилирования для системы LuNA (принято в печать).

Заключение

В результате работы:

1. Сформулированы требования к подсистеме профилирования в системе программирования LuNA.
2. В соответствии с требованиями разработана архитектура подсистемы профилирования.
3. Реализован и протестирован прототип подсистемы, реализующий разработанную архитектуру.

Поставленные задачи выполнены, цель работы достигнута.

Заключение

Защищаемые положения:

1. Архитектура подсистемы профилирования для системы фрагментированного программирования LuNA.
2. Экспериментальное исследование работоспособности прототипа подсистемы профилирования для системы фрагментированного программирования LuNA. Исследование показало, что использование прототипа позволило улучшить производительность исполнительной системы при трассировке фрагментированных программ.

Дальнейшие планы

1. Расширение функциональности прототипа для встраивания в другие проекты системы программирования LuNA, использующие профилирование.
2. Реализация разработанных способов уменьшения накладных расходов на работу подсистемы профилирования:
   - возможности выбора необходимых модулей подсистемы профилирования при компиляции системы LuNA;
   - возможности включать модули при запуске системы программирования LuNA с помощью интерфейса командной строки.
3. Разработка способов уменьшения накладных расходов на поддержку многопоточности.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ, НГУ)

Факультет информационных технологий
Кафедра параллельных вычислений

Направление подготовки 09.03.01 Информатика и вычислительная техника
Направленность (профиль): Программная инженерия и компьютерные науки

РАЗРАБОТКА И РЕАЛИЗАЦИЯ ПОДСИСТЕМЫ ПРОФИЛИРОВАНИЯ ДЛЯ СИСТЕМЫ ФРАГМЕНТИРОВАННОГО ПРОГРАММИРОВАНИЯ LUNA

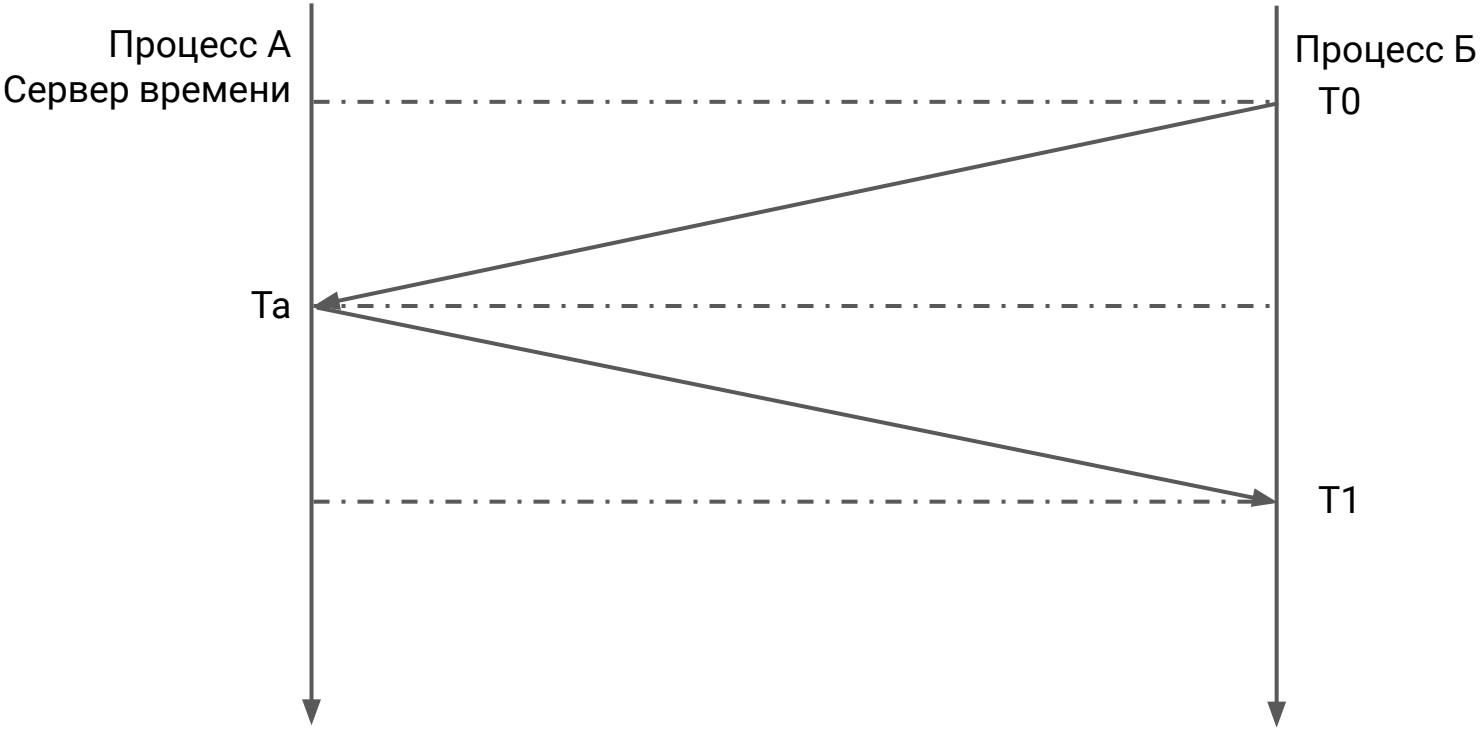
Выполнила: Абрамушкина Екатерина Сергеевна, студентка группы 19202 ФИТ НГУ
Научный руководитель: Маркова Валентина Петровна, д.т.н., доцент каф. ПВ ФИТ НГУ
Соруководитель: Киреев Сергей Евгеньевич, ст. преп. каф. ПВ ФИТ НГУ

Новосибирск 2023

23.06.2023

Научная новизна и практическая ценность

Научная новизна работы заключается в **разработке** подсистемы профилирования в системе фрагментированного программирования LuNA, которая позволит объединить существующие инструменты профилирования и предоставит возможности для расширения функциональности средств профилирования в системе LuNA.

Практическая ценность работы состоит в том, что разработанная подсистема позволит унифицировать процесс сбора информации о событиях, происходящих в процессе исполнения фрагментированных программ. Подсистема предоставит единый интерфейс для работы с собранными данными, для использования которого в работе также предложен универсальный и расширяемый формат профиля.

Алгоритм Кристиана синхронизации времени