



**Центр**  
компетенций  
и обучения

XV Международная конференция  
Научный сервис в сети Интернет: все  
границы параллелизма  
23-28 сентября 2013, Абрау-Дюрсо

ПРЕДПРИЯТИЕ ГОСКОРПОРАЦИИ «РОСАТОМ»

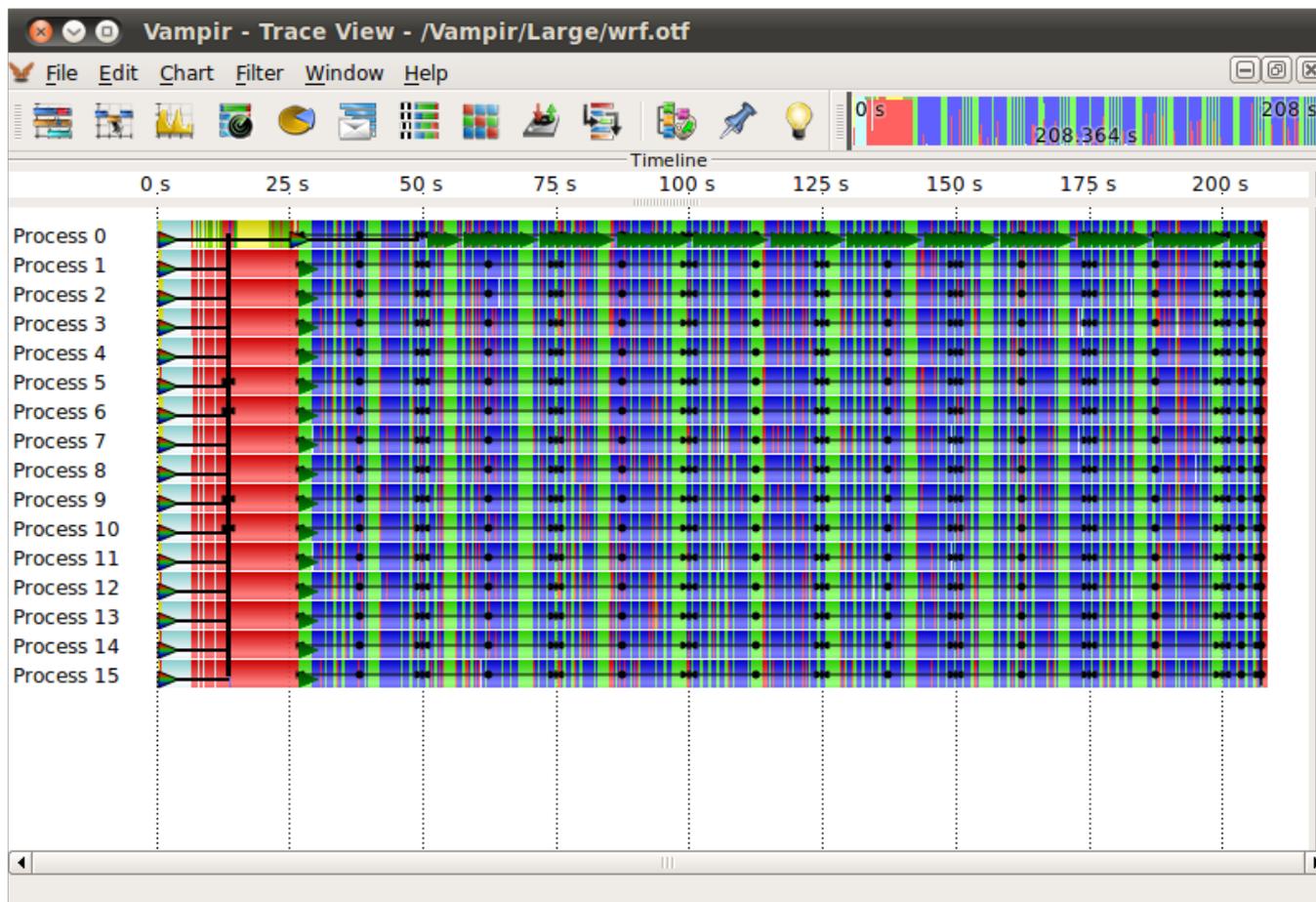
# **ХТФ – Масштабируемый формат хранения трассировочной информации**

Д.В. Ежов, А.В. Ершов, А.В. Огородников,  
ООО «Центр Компетенций и Обучения», г. Саров,  
А.Д. Модяев, НИЯУ МИФИ, г. Москва  
e-mail: [info@compcenter.org](mailto:info@compcenter.org)

# Анализ параллельных MPI приложений

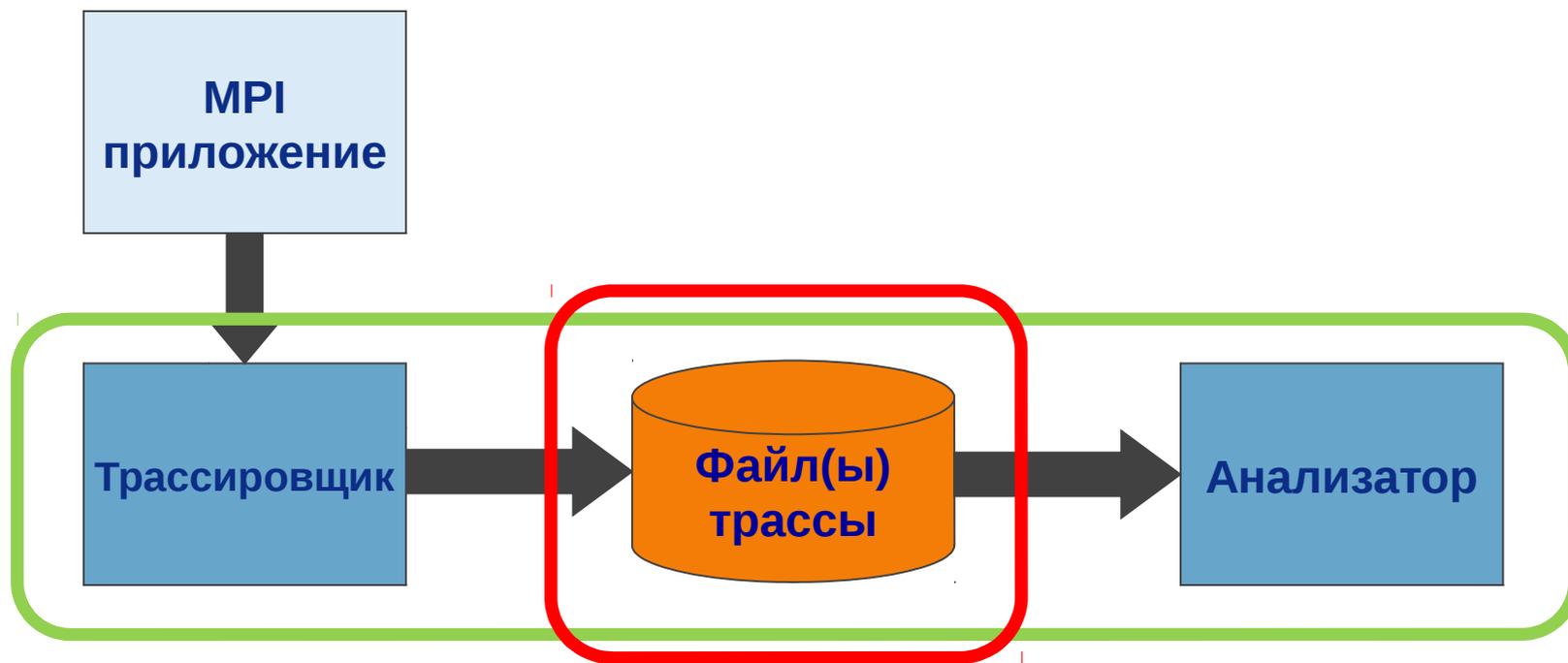
- Статистика
  - Общая количественная информация о приложении
- Профилирование
  - Зависимость достоверности информации от частоты профилирования
- Трассировка
  - Регистрация 100% событий

# Пример графического отображения трассы – Vampir\*

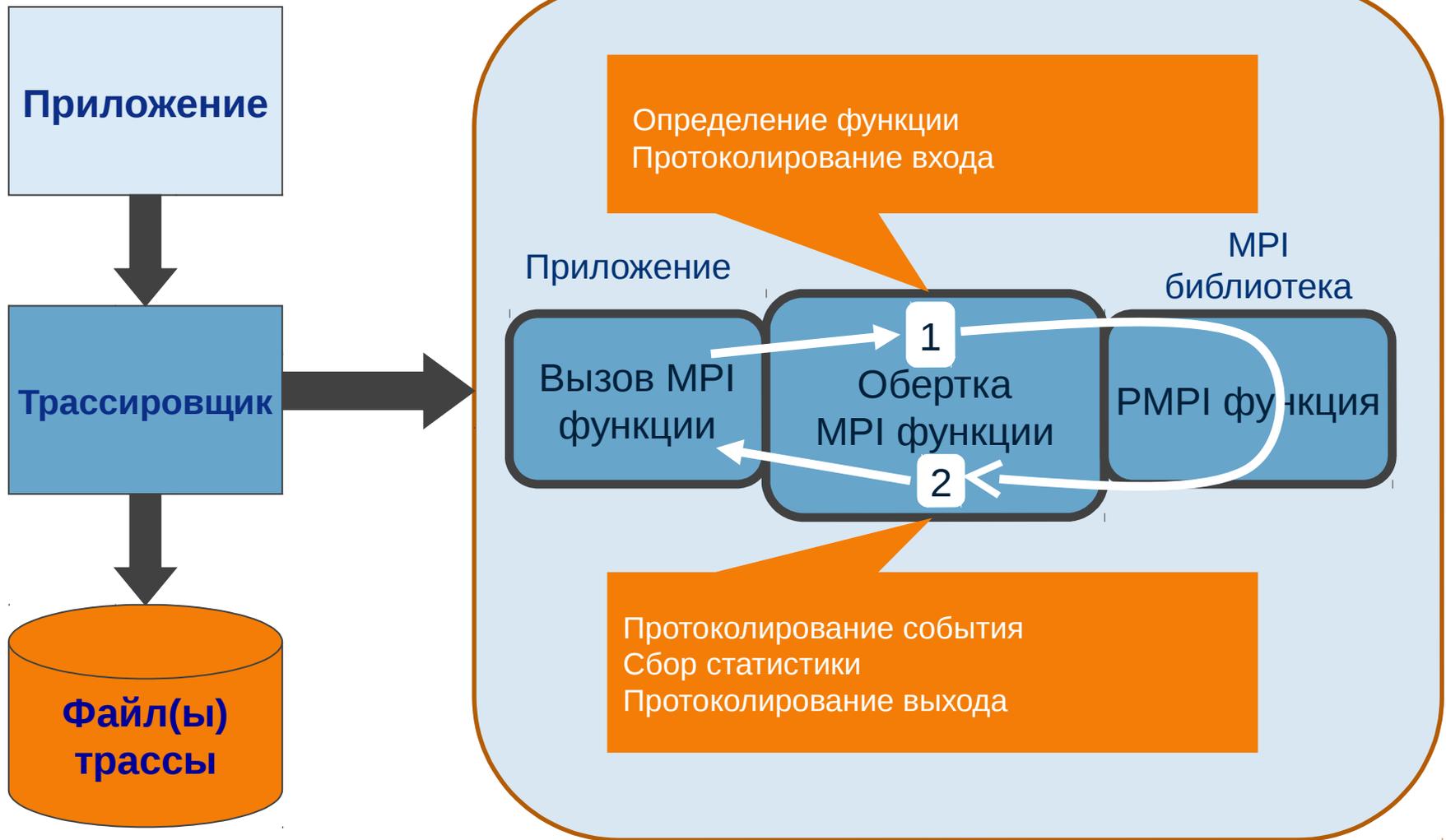


\* разработка компании GWT-TUD GmbH, [www.vampir.eu](http://www.vampir.eu)

# Схема взаимодействия инструментов анализа



# Схема работы трассировщика



# Проблемы анализа трасс



# Масштабируемый анализ

- Ослабление зависимости времени обработки трассы от ее размера
- Исключение необходимости чтения всей трассы или крупных блоков для поиска конкретного события
- Снижение затрат на постобработку трассы
- Новый высокомасштабируемый формат ХТФ позволяет:
  - Осуществлять прямой доступ к данным
  - Реализовать новые методики обработки
  - Реализовать новые методики визуализации
  - Осуществлять кеширование загруженных данных

# ХТФ: масштабируемый иерархический формат

- Иерархический контейнер данных
  - Страничное распределение памяти в файлах
  - Индексированные массивы данных
  - Прозрачное кеширование
- Доступ к данным
  - Индексный: поиск записи по ключу (ключам)
  - Прямой: доступ по ссылке
  - Последовательный: чтение интервала записей

# Данные и индексы

- Данные
  - Определения объектов трассы: классы объектов, события, идентификаторы, коммутаторы и т.д.
  - События: функции, двухточечные и коллективные операции и т.д.
  - Статистика
- Первичные индексы
  - Номер процесса P
  - Объединенный TN
    - Время события T
    - Номер события N
- Вторичные индексы
  - Классы объектов

# Индексы: В-деревья

- Сильно ветвящееся сбалансированное дерево типа B-tree
- Два вида узлов
  - Индексные: ключи + ссылка к странице
  - Листья: ключи + данные
- Высота дерева  $H$  для листьев одинакова
- Заполненность страниц 100%
  - Кроме страниц «правого края» дерева
  - Возникает автоматически при наполнении

# Время обработки: $O(\log N)$

- Высота дерева
- $H \approx 1 + \log S(N/Z) = (1 - \log SZ) + \log SN \geq 1 + \log SN$ 
  - $N$  = общее число записей массива
  - $S$  = число записей на индексной странице
  - $Z$  = число записей на листе
- Определяет число загрузок страниц при поиске
- Число сравнений ключей в памяти
- $C \leq \log_2(N/Z) + \log_2 Z = \log_2 N$   
(для бинарного поиска на странице в памяти)

# Пример

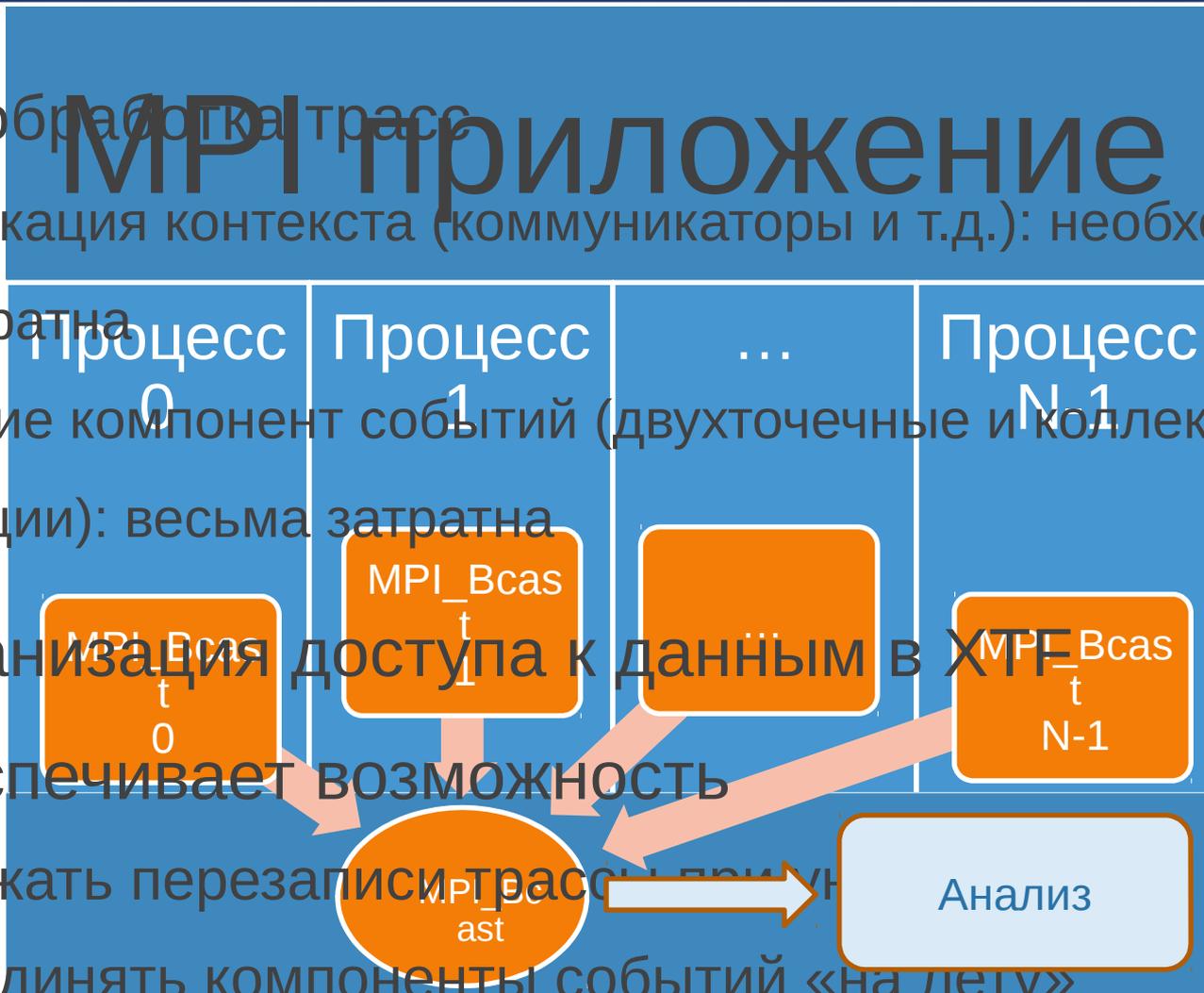
- Объемы
  - Массив: 100Gb, страница: 4Kb
  - Индексная запись: 24b, запись данных: 64b
- Размеры
  - $N = 1677721600$ ,  $S = 170$ ,  $Z = 64$ ,  $H = 5$
  - Количество страниц на каждом уровне:  
1, 6, 908, 154203, 26214400
    - Индекс: 155118 страниц ( $\approx 0.6\%$  общего объема)
  - Трудоемкость поиска
    - Не более 5 загрузок страниц, реально: 3-4
    - Не более 31 сравнения ключей в памяти

# Методики обработки

- Прimitives доступа
  - Поиск по ключу
  - Определение числа записей на интервале [key1, key2]
  - Перемещение по массиву записей
- «Быстрые» алгоритмы для
  - Приближенных графических представлений
  - Приближенных статистических характеристик
  - Интерактивной навигации по трассе

# Уменьшение затрат на постобработку

- Постобработка трассы
- Унификация контекста (коммуникаторы и т.д.): необходима, но не затратна
- Слияние компонент событий (двухточечные и коллективные операции): весьма затратна
- Организация доступа к данным в ХТФ обеспечивает возможность
- Избежать перезаписи трассы при анализе
- Объединять компоненты событий «на лету»



# Дальнейшее развитие

- Гибкое управление
  - Размером страницы и кеша страниц
  - Возможностью компрессии данных
    - Внешняя: на уровне файла
    - Внутренняя: на уровне страницы/записей
- Предоставление информации
  - Об активных вызовах функций
  - Об активных операциях обмена
  - О статистике на интервалах

**Спасибо за внимание.  
Вопросы?**



---

**ООО «Центр компетенций и обучения»**  
**[info@compcenter.org](mailto:info@compcenter.org)**  
**(83130) 6-70-67**