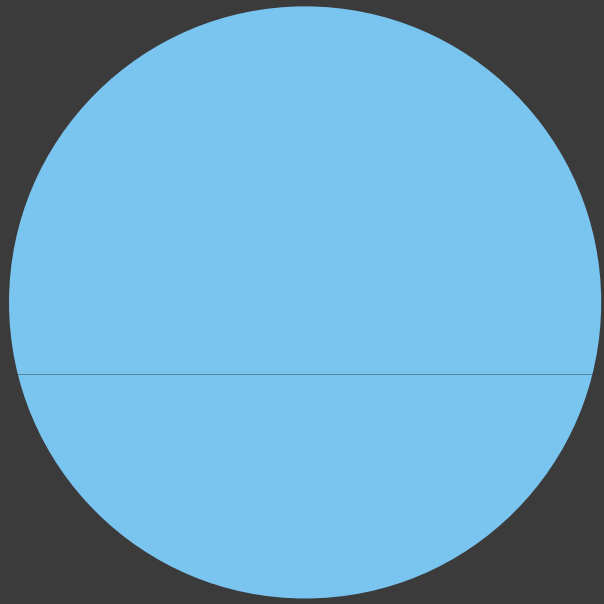
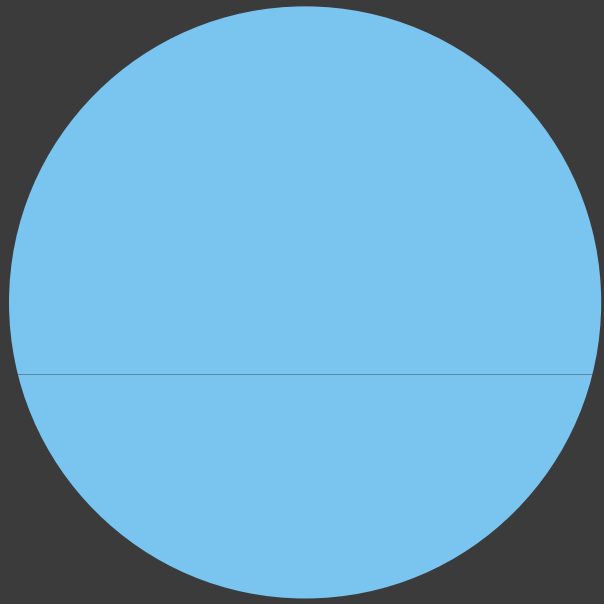
Что такое ТАС?



С.А.НЕМНЮГИН
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
2013

Тестироание



Академия Intel

Intuit.ru

Введение в программирование на кластерах
Оптимизация приложений с использованием библиотеки Intel Math Kernel Library. Уровень 1
Введение в программирование на Intel Cilk Plus
Введение в программирование на кластерах

Общая характеристика Intel® Trace Analyzer and Collector

Intel® Trace Analyzer and Collector

Intel® Trace Collector

Инструмент трассировки параллельных приложений, использующих технологию Message Passing Interface (MPI), приложений, работающих с общей памятью, а также обычных (не-MPI) приложений. Построен на основе Vampirtrace.

Intel® Trace Analyzer

Инструмент анализа и визуализации результатов трассировки параллельных приложений, использующих технологию Message Passing Interface (MPI). Графический интерфейс. Разнообразные виды анализа и графического представления его результатов.

Многоплатформенность

Microsoft Windows и Linux. Совместимость с конкретной платформой можно проверить на сайте http://www.intel.com/software/products/cluster.

Входят в состав Intel® Cluster Studio XE

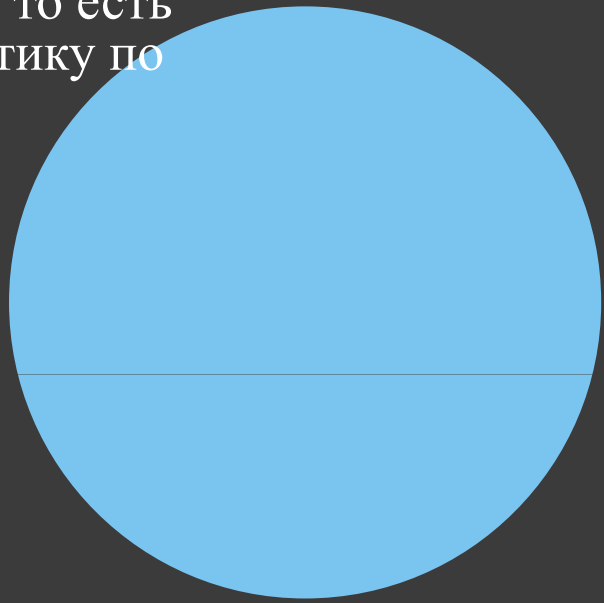
Трассировка

Intel® Trace Collector перехватывает вызовы MPI-функций. Для каждой функции MPI имеется своя «обёртка» (wrapper), которая позволяет выполнять дополнительные проверки, не предусмотренные в стандартных реализациях MPI.

Intel® Trace Collector

Intel® Trace Analyzer and Collector

Трассировка выполняется с помощью инструментовки приложения, то есть внедрения в него обращений к функциям, которые собирают статистику по различным событиям.



Виды инструментовки:

- бинарная;
- компиляторная;
- на уровне исходного кода.

При инструментовке на уровне исходного кода используется заголовочный файл VT.h (C/C++) или включаемый файл VT.inc (Fortran).

При инструментовке любого вида используются библиотеки Intel® Trace Collector.

Библиотеки Intel® Trace Collector

Библиотека	**Назначение**
libVTnull	«Заглушка»
libVT	Трассировка MPI-приложений и SHMEM-приложений
libVTfs	Безопасная трассировка MPI-приложений и SHMEM-приложений (результаты трассировки сохраняются даже после аварийного завершения приложения)
libVTmc	Проверка корректности
libVTcs	Трассировка распределённых приложений
VT_sample	Автоматическая трассировка счётчиков с PAPI и getrusage

Утилиты Intel® Trace Collector

Утилита	**Назначение**
stftool	Преобразование файлов с результатами трассировки
xstftool/expandvtlog.pl	Преобразование трассировочных файлов в читаемый формат
itcpin	Трассировка бинарных файлов без перекомпиляции

Сборка исполняемого файла для трассировки MPI-приложения

Linux

```
mpicc app.o -L$VT_LIB_DIR -lVT $VT_ADD_LIBS -o app
mpif77 app.o -L$VT_LIB_DIR -lVT $VT_ADD_LIBS -o app
```

Microsoft* Windows*

```
mpiicc app.obj /link /LIBPATH:%VT_LIB_DIR% VT.lib
mpiifort app.obj /link /LIBPATH:%VT_LIB_DIR% VT.lib
```

Идентификация событий в файле трассировки

Стандартные коммуникаторы
COMM_WORLD для MPI_COMM_WORLD
COMM_SELF для MPI_COMM_SELF

Пользовательские коммуникаторы - имя формируется из префикса и имени «старого» коммуникатора.

Для MPI_Comm_create() префикс CREATE

Для MPI_Comm_dup() префикс DUP

Для MPI_Comm_split() префикс SPLIT

Для MPI_Cart_sub() префикс CART_SUB

Для MPI_Cart_create() префикс CART_CREATE

Для MPI_Graph_create() префикс GRAPH_CREATE

Для MPI_Intercomm_merge() префикс MERGE (<oldname 1>/<oldname 2>)

варийное завершение MPI-приложения приводит к потере результатов трассировки, если спользуется библиотека libVT.

варийное завершение MPI-приложения НЕ приводит к потере результатов трассировки, если спользуется библиотека libVTfs. В этом случае при аварийном завершении приложения его роцессы «замораживаются» до того момента, когда на диске будет сохранён файл трассировки.

иксируются события:

игналы – внутренние (ошибки сегментации, ошибки операций с плавающей точкой) и нешние (SIGINT, SIGTERM). SIGKILL не фиксируется.

реждевременное завершение – один или несколько процессов завершились до вызова PI_Finalize.

шибки MPI – ошибки обменов, неправильно заданные параметры функций MPI.

локировки – фиксируются, если в течение некоторого времени процессы простаивают аходятся в состоянии вызова одной MPI-функции). Время простоя задаётся с помощью EADLOCK-TIMEOUT.

Ошибки параллельного ввода-вывода фиксируются только в среде ОС Linux.

Intel® Trace Collector позволяет выполнять трассировку односторонних обменов.

Intel® Trace Collector позволяет выполнять трассировку приложений, работающих с общей памятью.

Intel® Trace Collector позволяет выполнять трассировку последовательных приложений (библиотека libVTcs). Для трассировки требуются вызовы VT_initialize() и VT_finalize().

«Фолдинг» (сокрытие информации) позволяет уменьшить количество трассировочной информации.

Бинарная инструментовка

`itcpin` выполняет подстановку библиотек ITC, их инициализацию, запись событий входа в функции и выхода из них.

```
itcpin [<ключи ITC>] -- <командная строка запуска приложения>
```

Если ключи не указаны, утилита выполняет проверку возможности инструментовки файла и определение способа такой инструментовки.

Ключ `--list` позволяет определить функции в исполняемом файле. Результат записывается в формате:

```
<имя бинарного файла>:<имя исходного файла>:<имя функции>
```

Трассировка

Ключ `--run` запускает приложение в режиме сбора статистики. По умолчанию, если в исполняемом файле содержатся вызовы MPI-функций, подключается библиотека libVT.

Ключ `--insert` используется для подключения определённой библиотеки.

На всех вычислительных узлах Collector должен быть установлен в каталогах с одинаковым маршрутным именем.

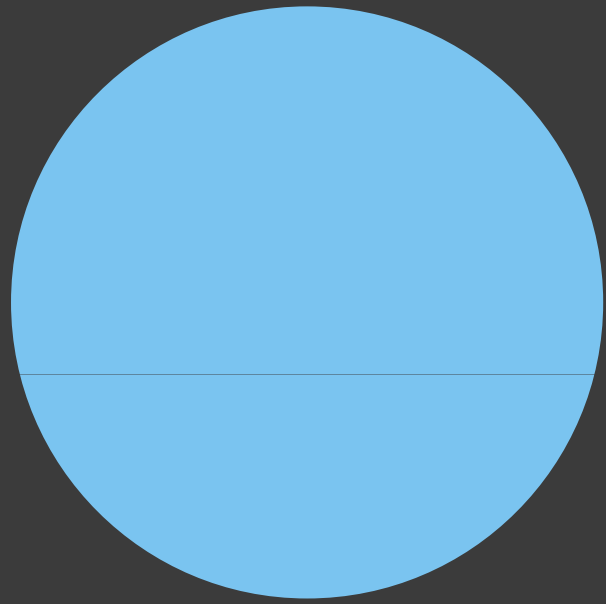
Ключ `--profile` используется для профилирования функций.

атистика сохраняется в **STF-**файле.

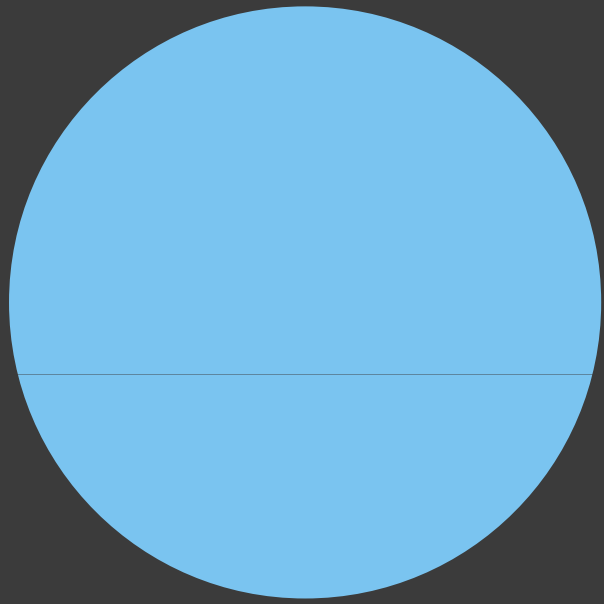
ждая строка содержит:
Поток или процесс.
Функция.
Получатель сообщения.
Размер сообщения.
Количество процессов.

обирается следующая статистика:
Количество обменов.
Минимальное время выполнения, исключая время вызываемых функций.
Максимальное время выполнения, исключая время вызываемых функций.
Суммарное время выполнения, исключая время вызываемых функций.
Минимальное время выполнения, включая время вызываемых функций.
Максимальное время выполнения, включая время вызываемых функций.
Суммарное время выполнения, включая время вызываемых функций.

поле получателя сообщения записывается:
`0xffffffff` для операций с файлами;
`0xfffffffe` для коллективных операций.



Счетчики



Сбор данных о производительности процессора

PAPI (Performance Application Programming Interface) – интерфейс, позволяющий собирать статистику с аппаратных счётчиков и системную статистику.

Поддержка PAPI реализована поверх ITC – в файле VT_sample.c.

Сбор статистики с аппаратных счётчиков включается с помощью опции конфигурации:

```
COUNTER <имя счётчика> ON
```

В имени счётчика допускается использование метасимвола *.

Системные счётчики (список неполон)

Название счётчика	**Единицы измерения**	**Собираемая информация**
RU_UTIME	Сек.	Время в режиме задачи
RU_STIME	Сек.	Время в режиме ядра
RU_MAXRSS	Байты	Максимальный размер резидентной части
RU_IXRSS	Байты	Суммарный размер разделяемой памяти
RU_MAJFLT	#	Ошибки страниц
RU_NSWAP	#	Выгрузки
RU_INBLOCK	#	Блочные операции ввода
RU_OUBLOCK	#	Блочные операции вывода
RU_MSGSND	#	Отправленные сообщения
RU_MSGRCV	#	Полученные сообщения
RU_NSIGNALS	#	Полученные сигналы

Системные счётчики (локальные, для узла)

Название счётчика	**Единицы измерения**	**Собираемая информация**
disk_io	кб/сек	Ввод-вывод на диск
net_io	кб/сек	Сетевой ввод-вывод (не включает транспортный уровень MPI)
cpu_ . . .	%	Средняя доля процессорного времени всех процессоров, проведенного в . . .
cpu_idle	%	. . . режиме простоя
cpu_sys	%	. . . режиме ядра
cpu_usr	%	. . . режиме задачи

Проверка корректности

Проверка корректности включает:

- проверку переносимости;
- проверку нарушений стандарта MPI, не приводящие к немедленным фатальным последствиям, но проявляющиеся при переходе на другие платформы или к другим реализациям MPI;
- проверку ошибок в среде исполнения.

Проверка корректности реализована в библиотеке libVTmc.

По умолчанию результаты проверки корректности не фиксируются. Включить запись можно установкой флага CHECK-TRACING в файле конфигурации.

Проверка корректности требует дополнительных ресурсов!

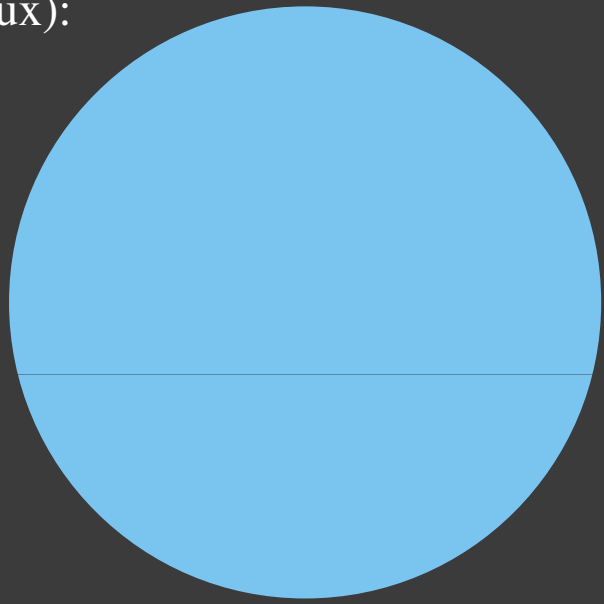
Проверка корректности реализована только для Intel® MPI Library!

Как это делается:

1 вариант.

При запуске использовать переменную окружения LD_PRELOAD (только Linux):

```
mpiexec -genv LD_PRELOAD libVTmc.so -n ...
```



2 вариант.

Бинарная инструментовка.

3 вариант.

Пересборка приложения.

4 вариант.

Использование ключа –check для mpiexec (только Intel® MPI).

Для того, чтобы добавить результат в трассировочный файл

Установить значение переменной окружения При запуске использовать переменную окружения VT_CHECK_TRACING, например:

```
mpiexec -genv LD_PRELOAD libVTmc.so -genv VT_CHECK_TRACING
on -n ...
```

Сигнатуры ошибок

Локальные ошибки

Сигнатура		Описание
LOCAL:EXIT:SIGNAL		Процесс остановлен «фатальным» сигналом
LOCAL:EXIT:BEFORE_MPI_FINALIZE		Процесс завершён без вызова MPI_Finalize()
LOCAL:MEMORY:OVERLAP		Несколько операций MPI используют одну область памяти
LOCAL:MEMORY:ILLEGAL_MODIFICATION		Некорректная модификация данных
LOCAL:MEMORY:INACCESSIBLE		Буфер недоступен
LOCAL:MEMORY:ILLEGAL_ACCESS		Некорректный доступ к памяти, уже используемой MPI
LOCAL:MEMORY:INITIALIZATION		Проверка распределённой памяти
LOCAL:REQUEST:ILLEGAL_CALL		Неправильная последовательность вызовов
LOCAL:REQUEST:NOT_FREED		Избыточное количество запросов или завершение программы с отложенными обменами
LOCAL:BUFFER:INSUFFICIENT_BUFFER		Недостаточно памяти для буферизованной отправки сообщения

Глобальные ошибки

Сигнатура ошибки	Описание
GLOBAL:MSG/COLLECTIVE:DATATYPE:MISMATCH	Несогласованность типов
GLOBAL:MSG/COLLECTIVE:DATA_TRANSMISSION_CORRUPTED	Модификация данных во время передачи
GLOBAL:MSG:PENDING	Завершение программы до приёма всех сообщений
GLOBAL:DEADLOCK:HARD	Цикл процессов, ожидающих друг друга
GLOBAL:DEADLOCK:NO_PROGRESS	Возможна блокировка
GLOBAL:COLLECTIVE:OPERATION_MISMATCH	Процессы участвуют в разных коллективных операция
GLOBAL:COLLECTIVE:SIZE_MISMATCH	Данных больше или меньше, чем должно быть
GLOBAL:COLLECTIVE:REDUCTION_OPERATION_MISMATCH	Ошибка в операции приведения
GLOBAL:COLLECTIVE:INVALID_PARAMETER	Неправильные параметры коллективной операции

Intel® Trace Analyzer

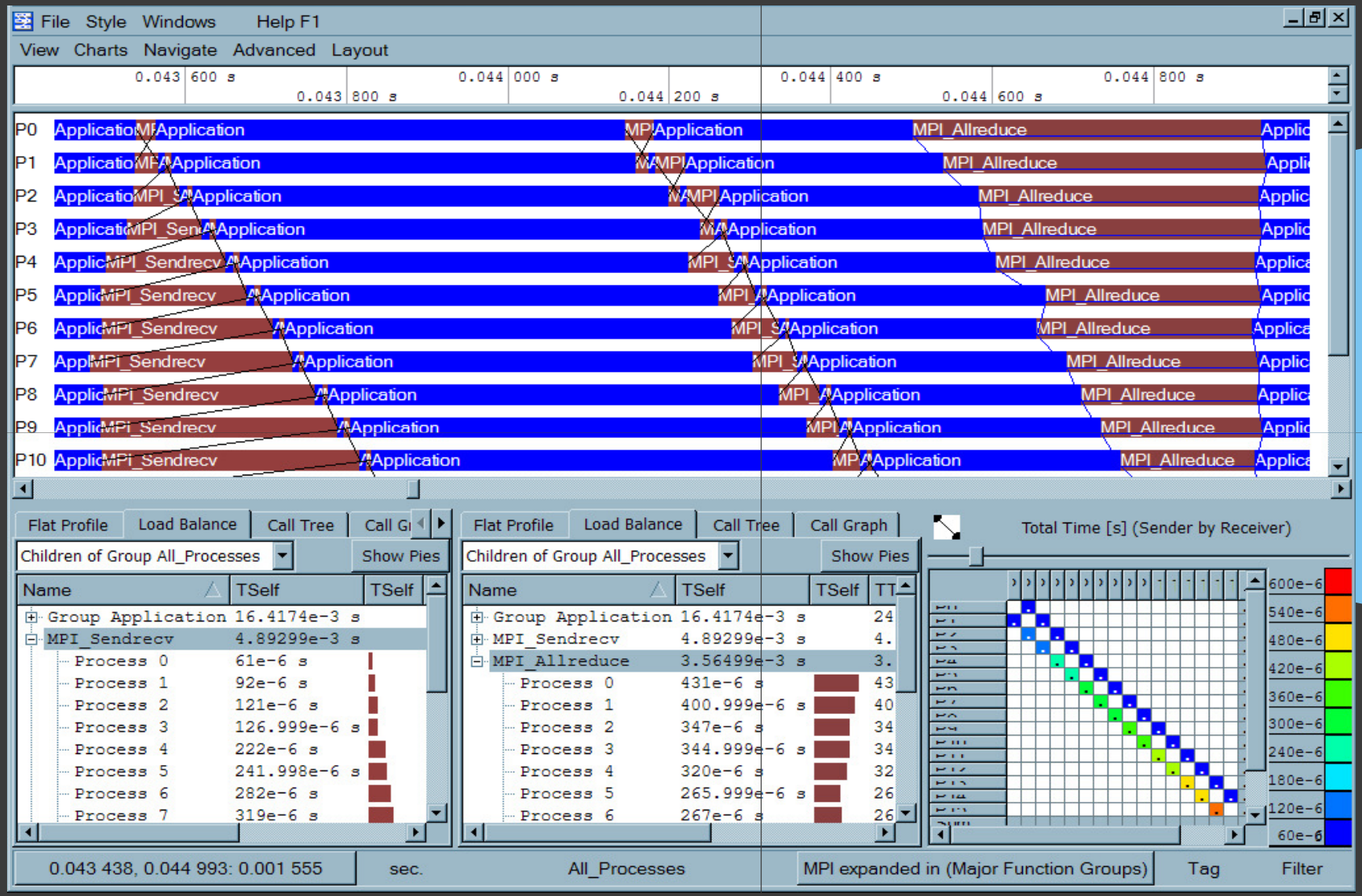
Запуск анализатора:

```
# traceanalyzer файл_трассировки.stf
```
(ОС Linux)

Start -> All Programs -> Intel Trace Analyzer (MS Windows)

Поддерживается как графический интерфейс, так и интерфейс командной строки.

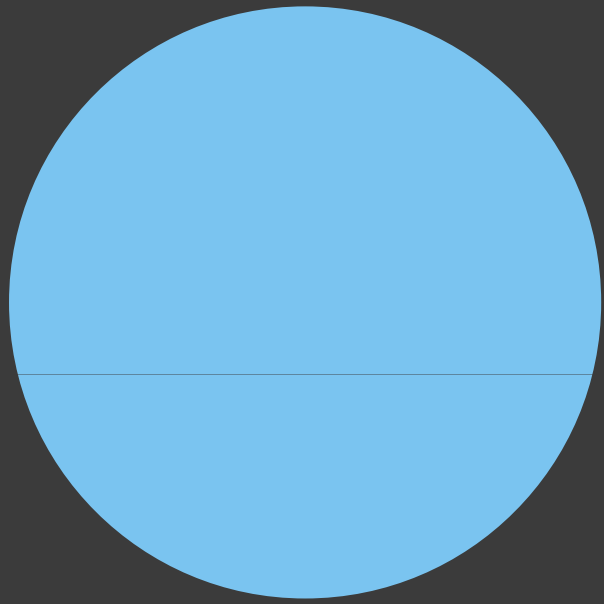
Пример вывода результатов анализа



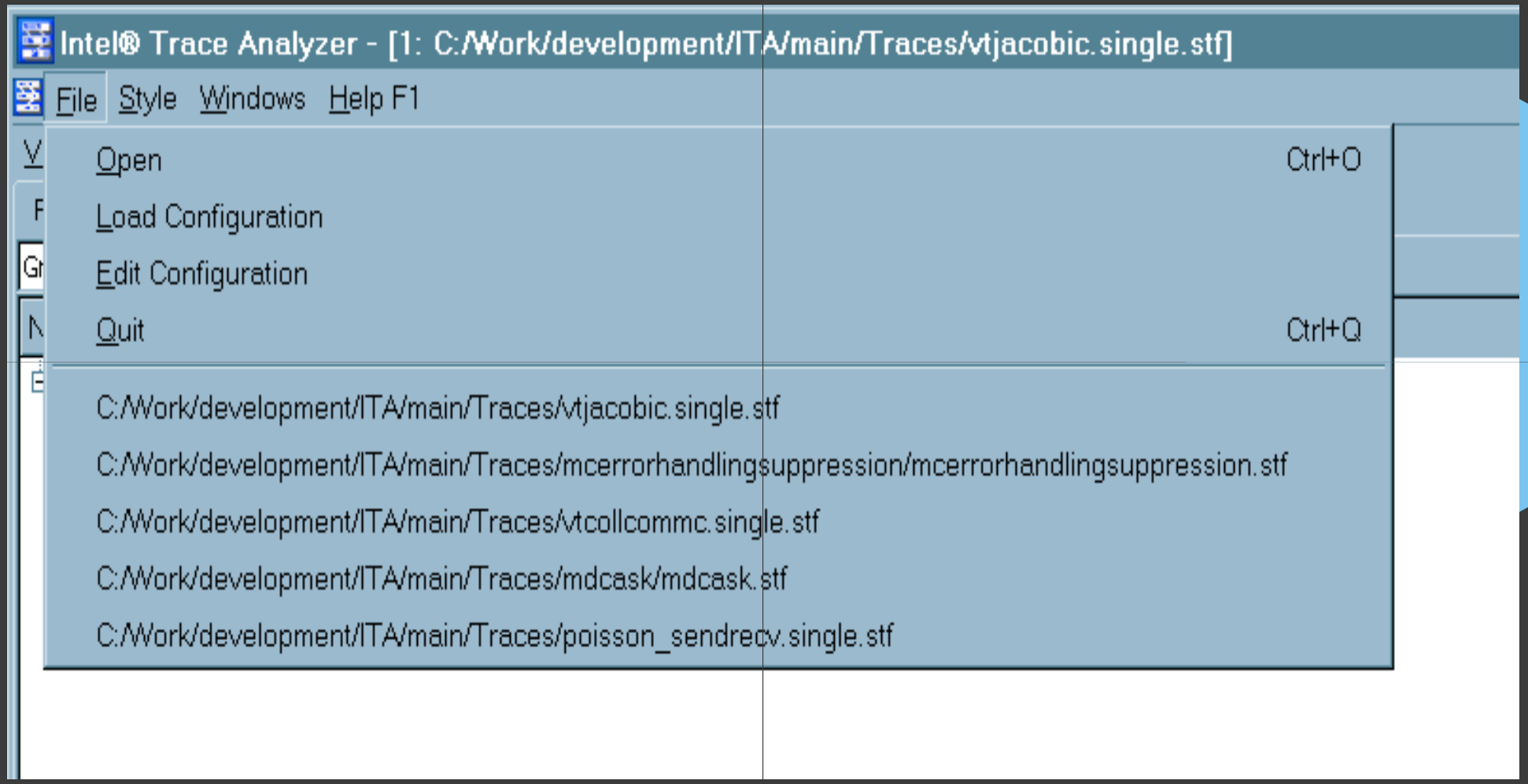
Графический интерфейс Intel® Trace Analyzer

Главное меню

1. File Menu
2. Style Menu
3. The Windows Menu
4. Help Menu

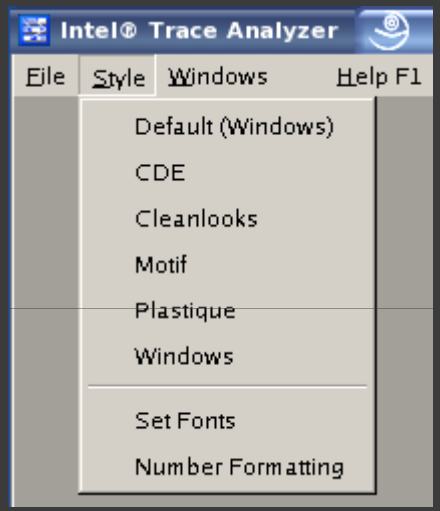


Меню работы с файлами

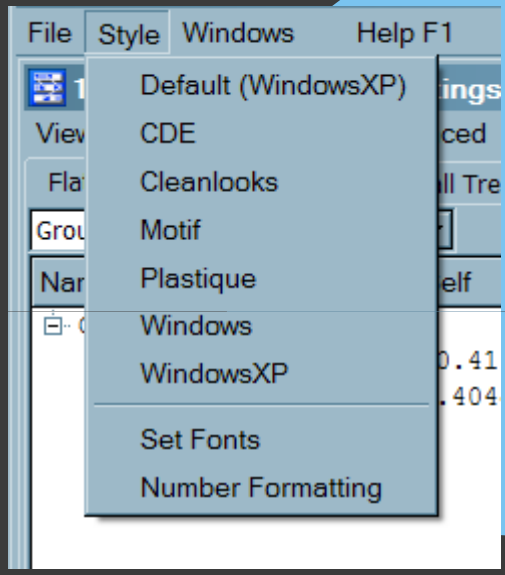


Меню стиля

Linux

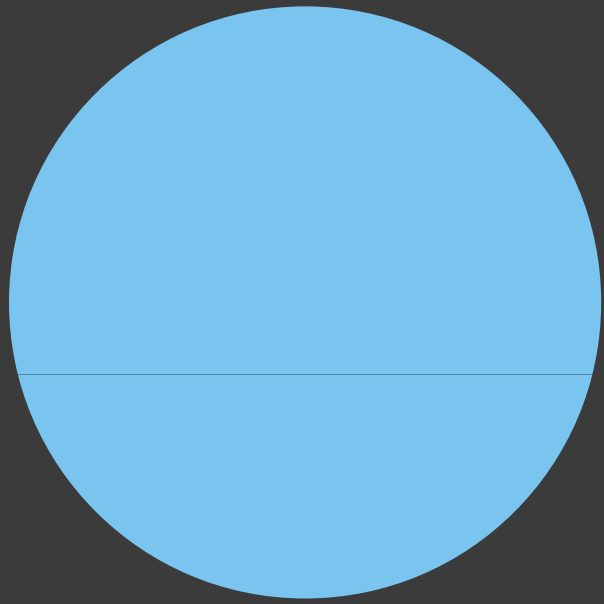


Windows

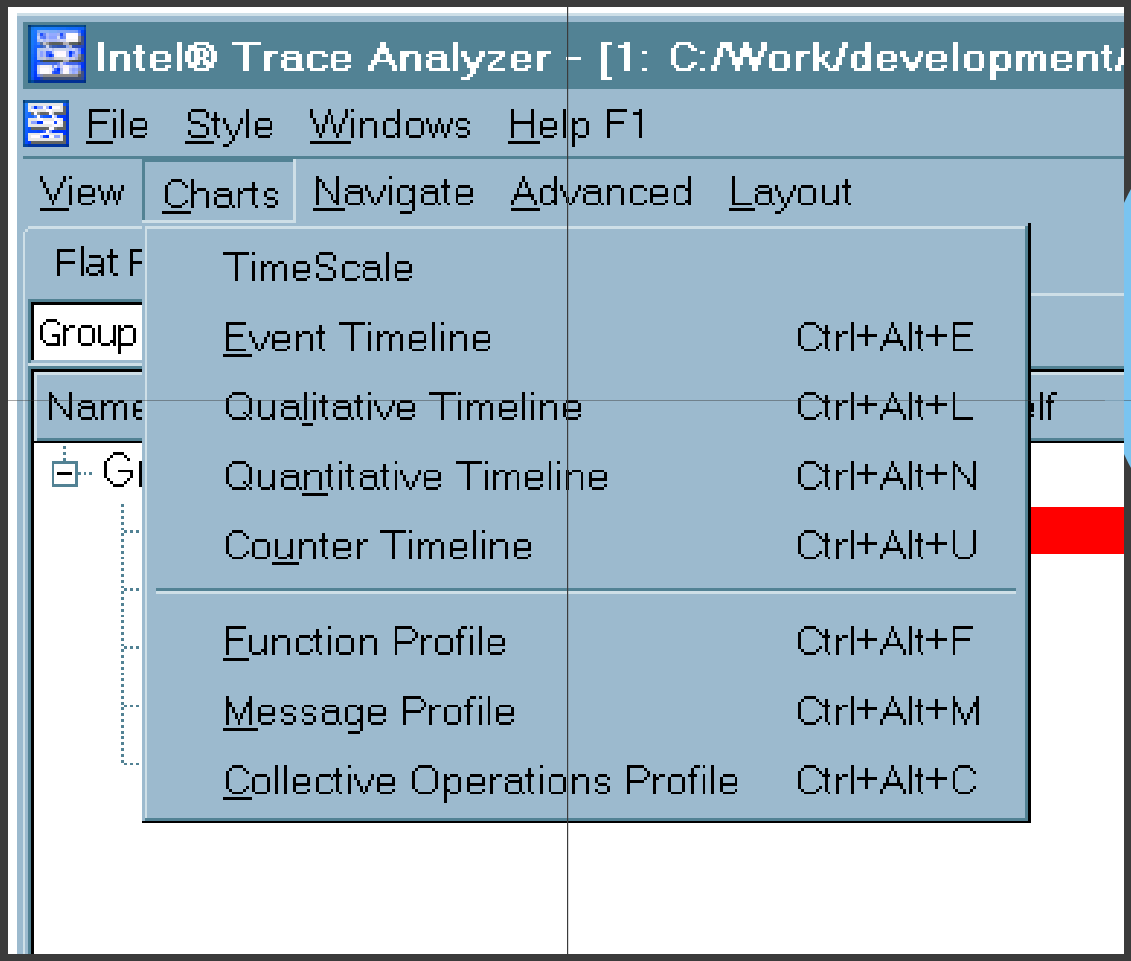


Меню окон просмотра

1. View Menu
2. Charts Menu
3. Navigate Menu
4. Advanced Menu
5. Layout Menu
6. Comparison Menu



Меню диаграмм

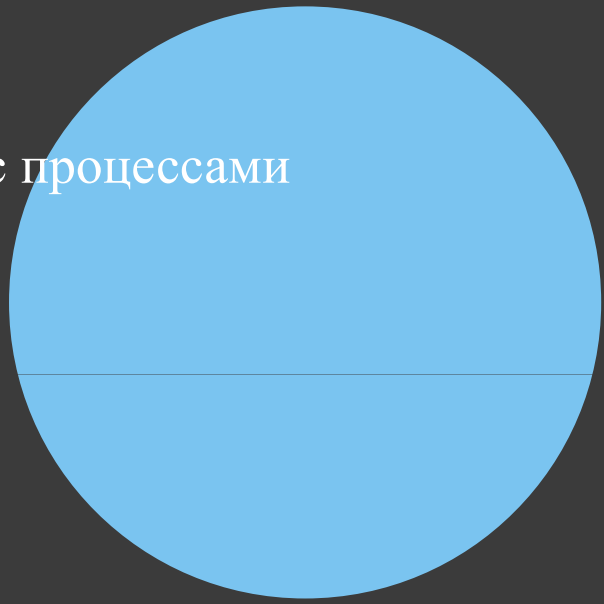


Time Scale

Временная шкала.

Event Timeline

Временная диаграмма, на которой отображаются события, связанные с процессами параллельной программы.



Qualitative Timeline

Временная диаграмма, на которой отображаются атрибуты событий.

Quantitative Timeline

Временная диаграмма, на которой отображается поведение параллельной программы.

Function Profile

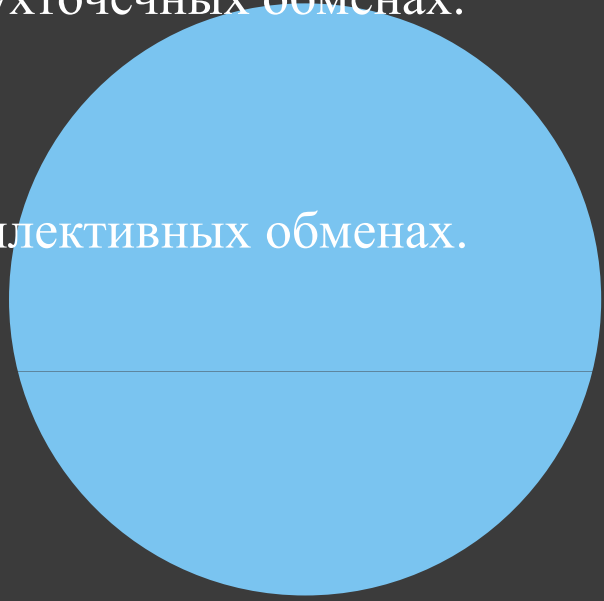
Диаграмма, на которой отображается информация, связанная с функциями программы.

Message Profile

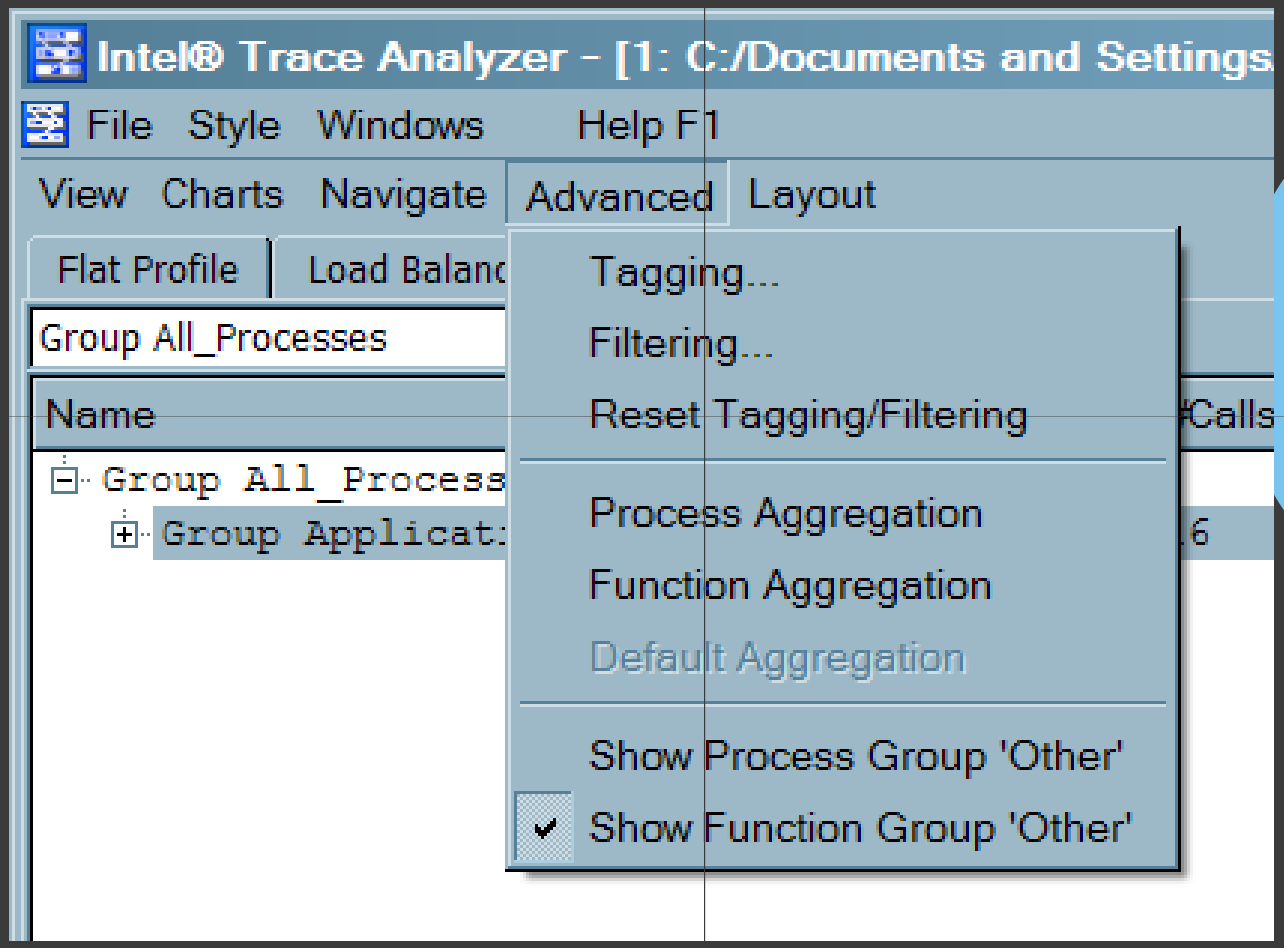
Диаграмма, на которой отображается статистическая информация о двухточечных обменах.

Collective Operations Profile

Диаграмма, на которой отображается статистическая информация о коллективных обменах.



The Advanced Menu



Tagging
Подсветка событий, удовлетворяющих условиям, заданным пользователем.

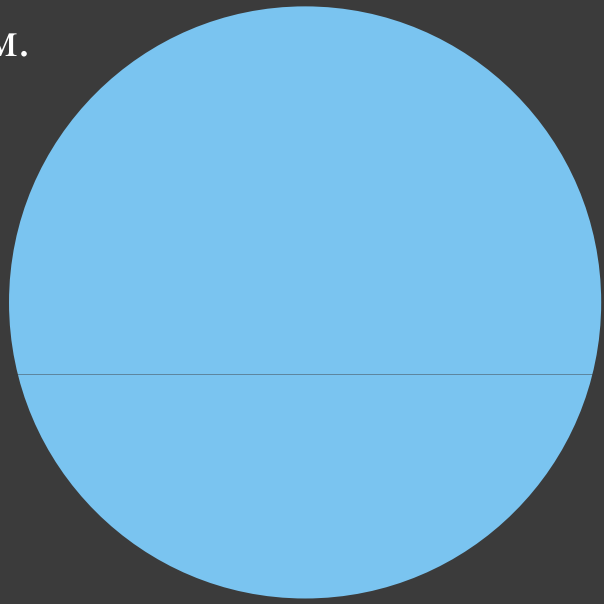
Filtering
Фильтрация событий, удовлетворяющих условиям, заданным пользователем.

Reset Tagging/Filtering
Откат операций маркировки и фильтрации.

Process Aggregation
Объединение результатов трассировки по группам процессов.

Default Aggregation
Установка параметров аггрегации по умолчанию.

Show Process Group “Other” и **Show Function Group “Other”**
Отображение результатов по процессам и функциям, не попавшим в объединение.



Диаграммы

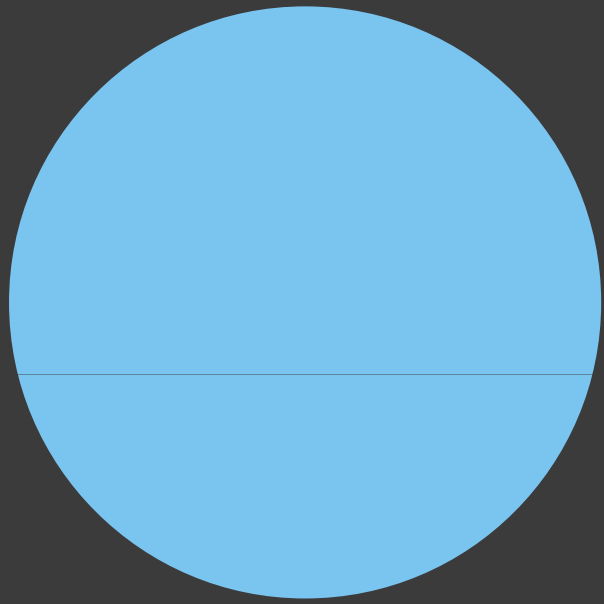
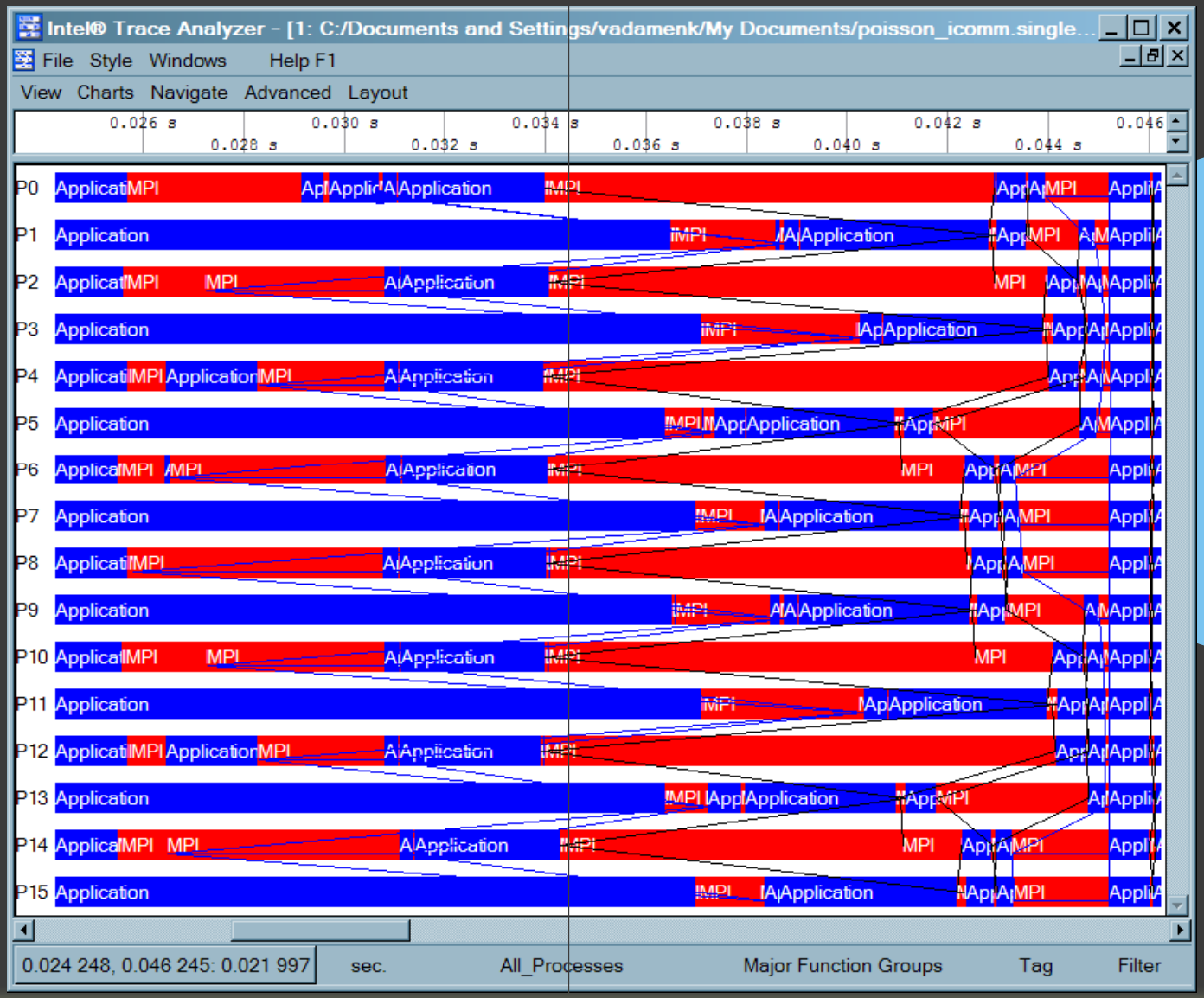


Диаграмма событий



Отображает активность индивидуальных процессов.

Горизонтальная ось – время.

Вертикальная ось – процесс.

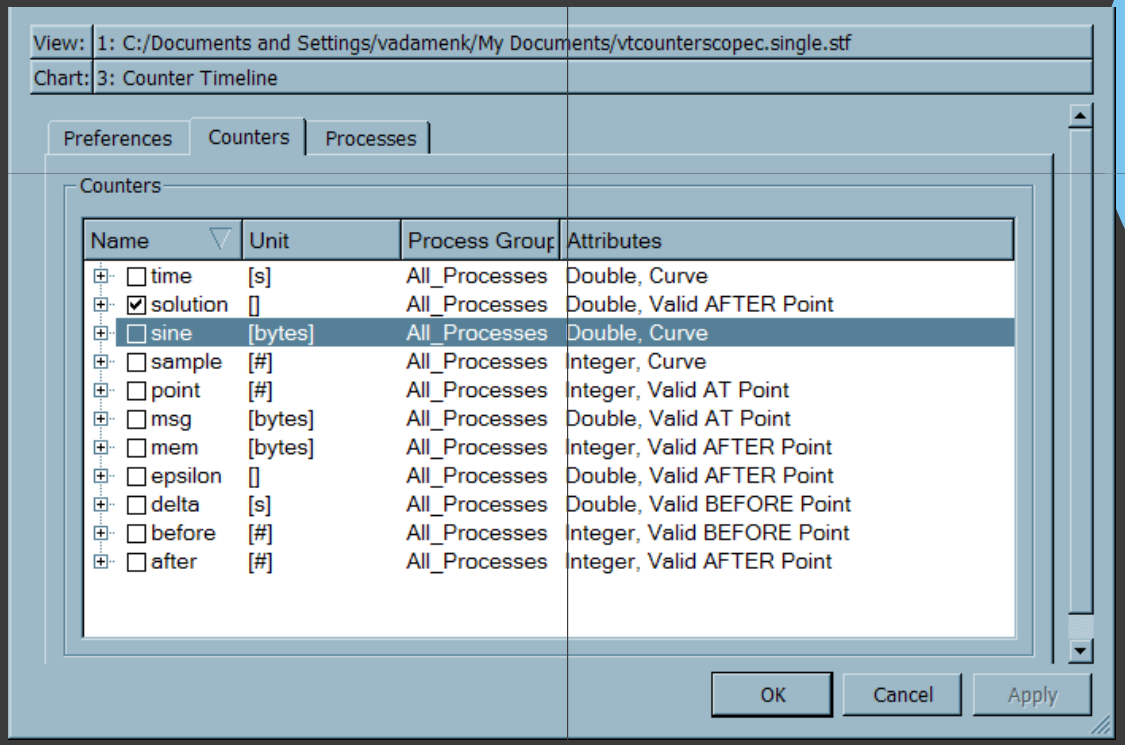
Чёрными линиями отображаются операции двухточечного обмена.

Синими линиями отображаются операции коллективного обмена.

Временной масштаб изменяется с помощью мыши.

Диаграмма счётчиков» отображает значения счётчиков, сохранённые в файле рассировки.

Пример определения отображаемых счётчиков:

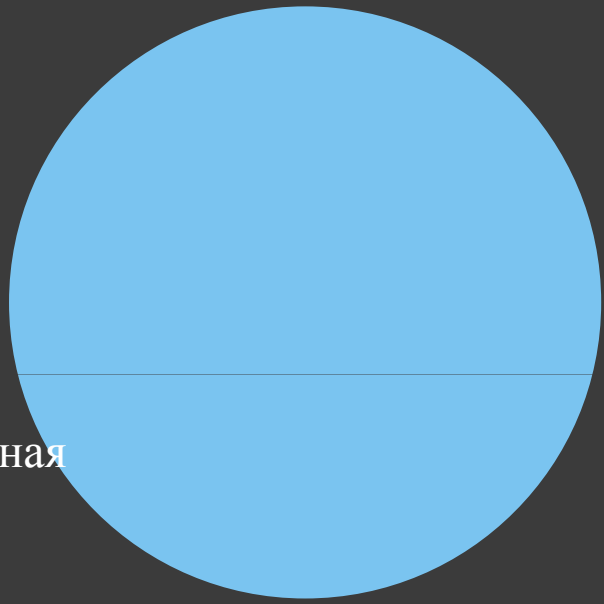


Профиль функций

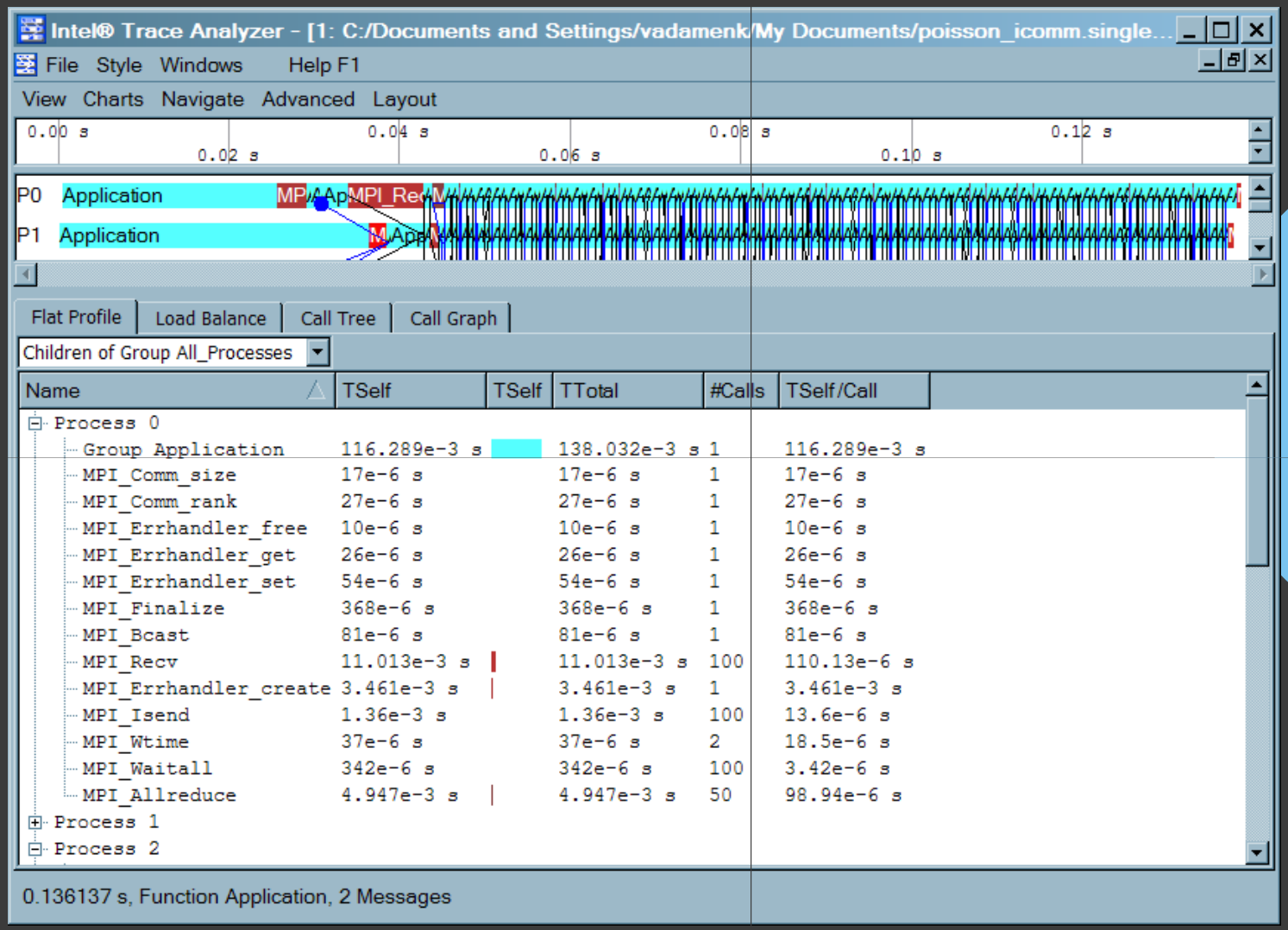
Диаграмма **«The Function Profile»** отображает детальную информацию о производительности.

Содержит вкладки:

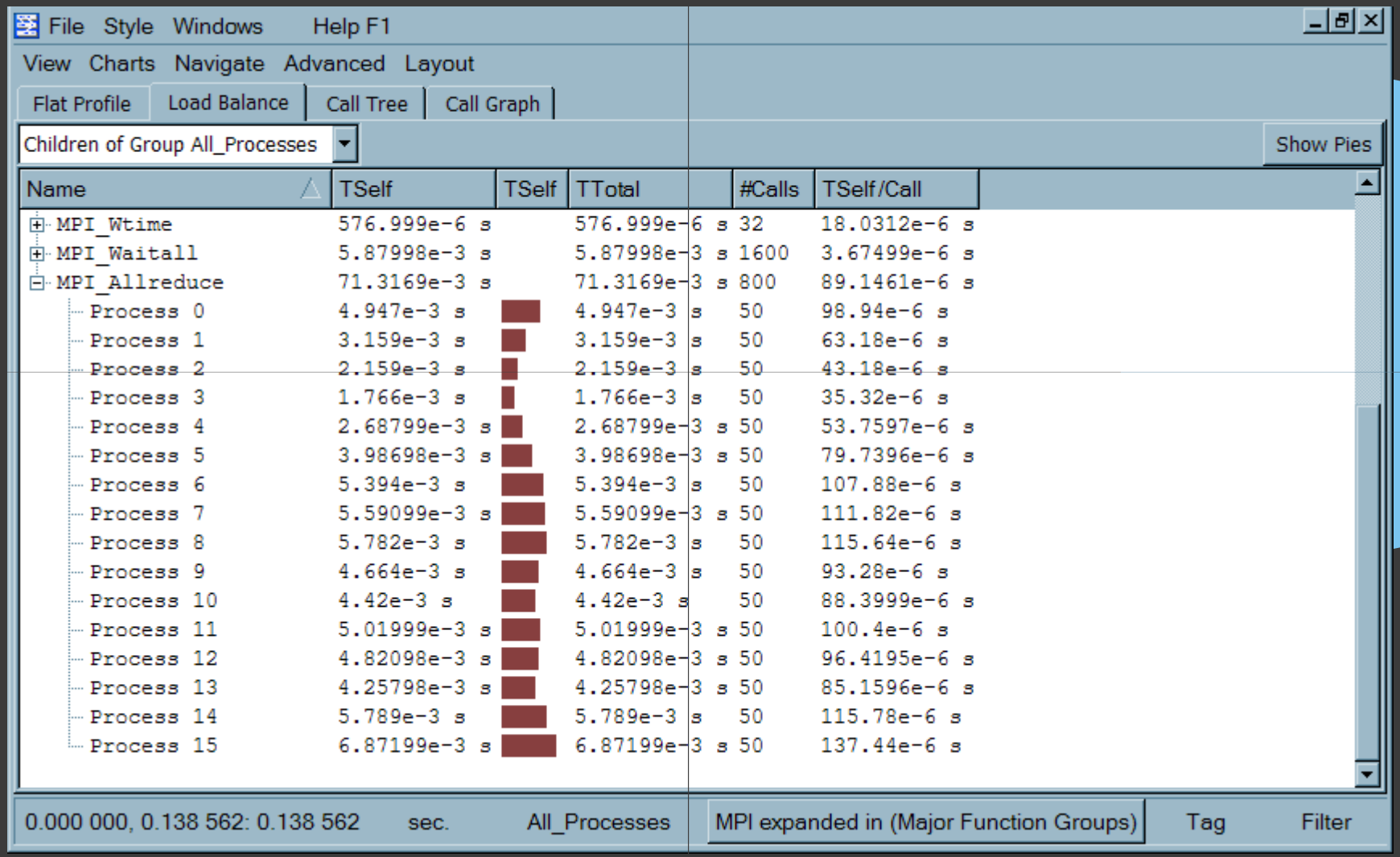
- Flat Profile – итоговая статистика по процессам.
- Load Balance – итоговая статистика по группам функций.
- Call Tree – последовательности вызовов.
- Call Graph – показывает небольшую часть графа вызовов (3 узла – центральная функция, вызывающая и вызываемая функции).



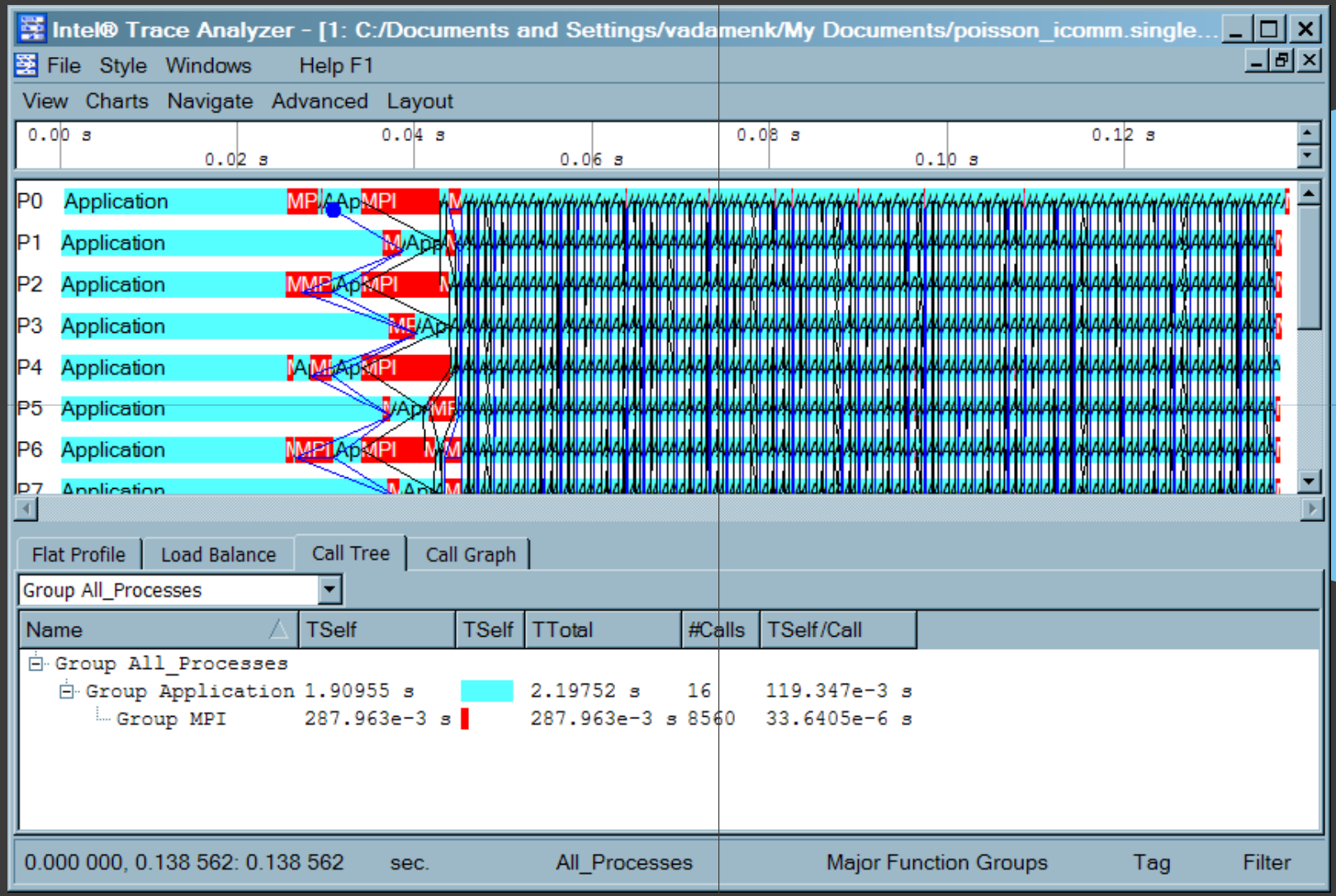
Flat Profile



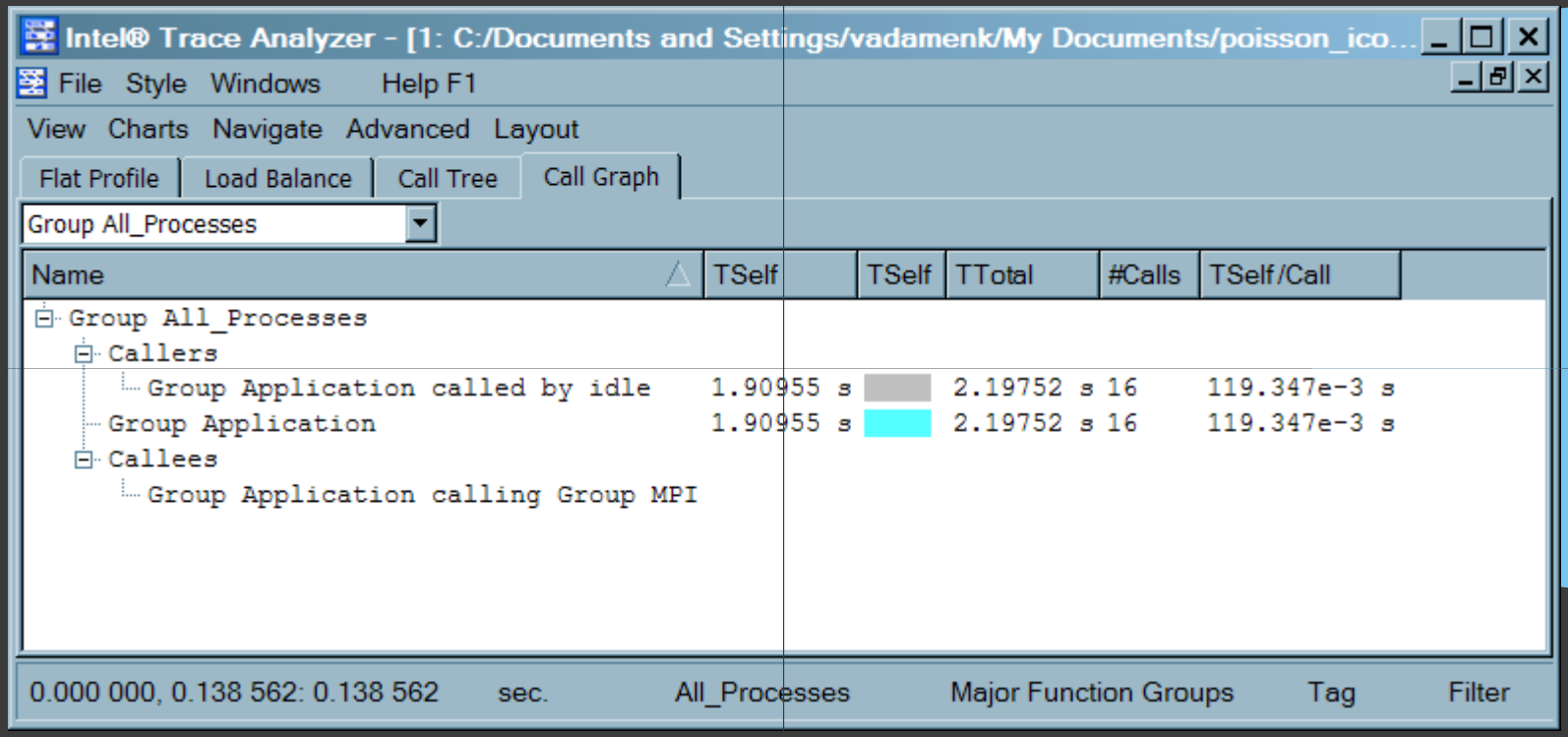
Load Balance



Call Tree


Intel® Trace Analyzer - [1: C:/Documents and Settings/vadamenk/My Documents/poisson_icomm.single...
File Style Windows Help F1
View Charts Navigate Advanced Layout
0.00 s
0.02 s
0.04 s
0.06 s
0.08 s
0.10 s
0.12 s
P0 Application
P1 Application
P2 Application
P3 Application
P4 Application
P5 Application
P6 Application
P7 Application
Flat Profile
Load Balance
Call Tree
Call Graph
Group All_Processes
Name
TSelf
TSelf
TTotal
#Calls
TSelf/Call
Group All_Processes
Group Application 1.90955 s 2.19752 s 16 119.347e-3 s
Group MPI 287.963e-3 s 287.963e-3 s 8560 33.6405e-6 s
0.000 000, 0.138 562: 0.138 562 sec.
All_Processes
Major Function Groups
Tag
Filter

Call Graph


Intel® Trace Analyzer - [1: C:/Documents and Settings/vadamenk/My Documents/poisson_ico...
File Style Windows Help F1
View Charts Navigate Advanced Layout
Flat Profile
Load Balance
Call Tree
Call Graph
Group All_Processes
Name
TSelf
TSelf
TTotal
#Calls
TSelf/Call
Group All_Processes
Callers
Group Application called by idle 1.90955 s 2.19752 s 16 119.347e-3 s
Group Application 1.90955 s 2.19752 s 16 119.347e-3 s
Callees
Group Application calling Group MPI
0.000 000, 0.138 562: 0.138 562 sec.
All_Processes
Major Function Groups
Tag
Filter

Профиль сообщений

Диаграмма **«The Message Profile»** отображает информацию об обменах в виде квадратной матрицы.

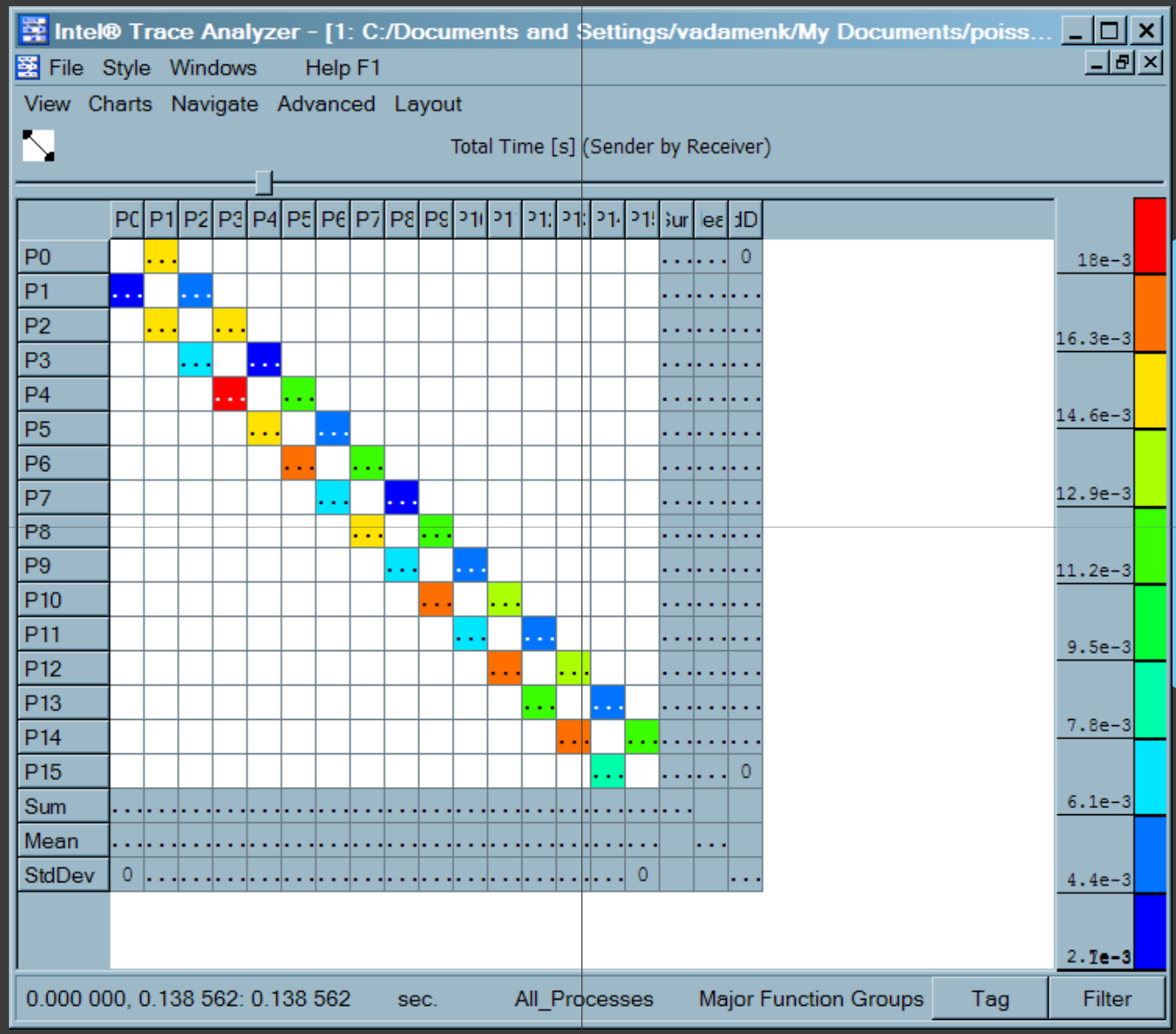
Строки матрицы соответствуют процессам-источникам сообщений.

Столбцы матрицы соответствуют процессам-приёмникам сообщений.

Каждая ячейка содержит суммарное время, затраченное на коммуникации между соответствующими процессами.

Приводятся также статистические характеристики по строкам и по столбцам.

Допускается группировка данных.


Intel® Trace Analyzer - [1: C:/Documents and Settings/vadamenk/My Documents/poiss...
File Style Windows Help F1
View Charts Navigate Advanced Layout
Total Time [s] (Sender by Receiver)
P0 P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 P15
Sum
Mean
StdDev
18e-3
16.3e-3
14.6e-3
12.9e-3
11.2e-3
9.5e-3
7.8e-3
6.1e-3
4.4e-3
2.1e-3
0.000 000, 0.138 562: 0.138 562 sec.
All_Processes
Major Function Groups
Tag
Filter

Профиль коллективных операций

Диаграмма **«The Collective Operations Profile»** отображает информацию о коллективных обменах в виде матрицы.

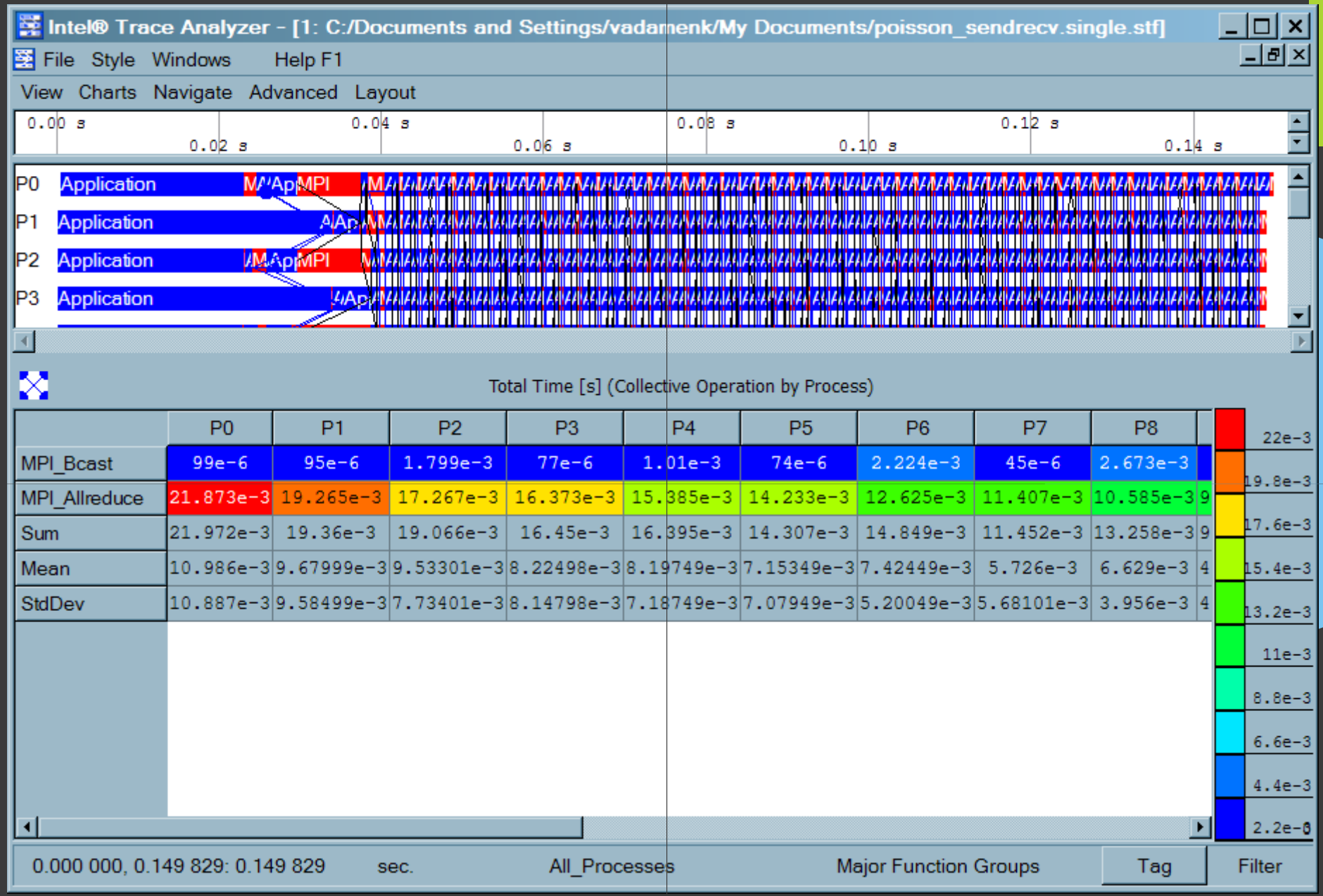
Строки матрицы соответствуют типам коллективных операций.

Столбцы матрицы соответствуют процессам-участникам обменов.

Каждая ячейка содержит суммарное время, затраченное на коммуникации между соответствующими процессами.

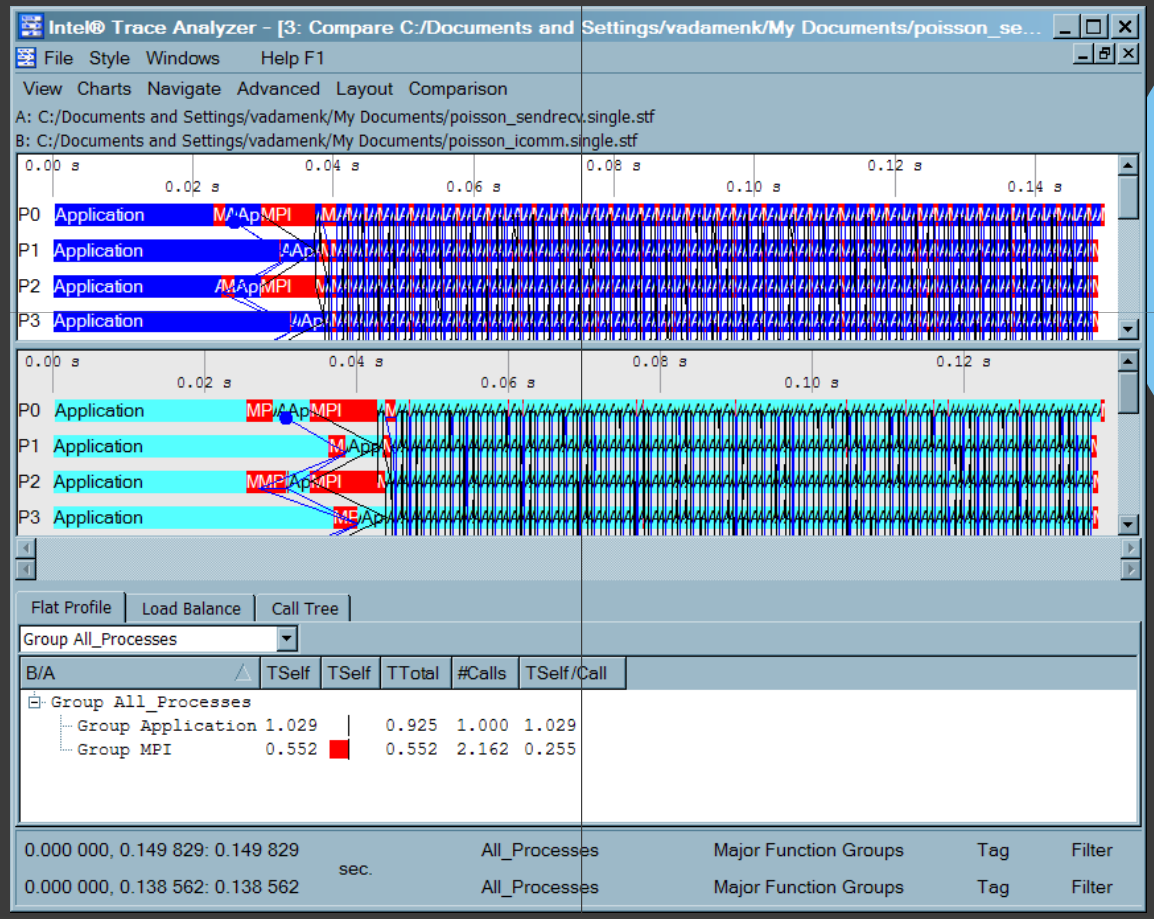
Приводятся также статистические характеристики по строкам и по столбцам.

Допускается группировка данных.


Intel® Trace Analyzer - [1: C:/Documents and Settings/vadamenk/My Documents/poisson_sendrecv.single.stf]
File Style Windows Help F1
View Charts Navigate Advanced Layout
0.00 s
0.02 s
0.04 s
0.06 s
0.08 s
0.10 s
0.12 s
0.14 s
P0 Application
P1 Application
P2 Application
P3 Application
MPI
Total Time [s] (Collective Operation by Process)
P0 P1 P2 P3 P4 P5 P6 P7 P8
MPI_Bcast 99e-6 95e-6 1.799e-3 77e-6 1.01e-3 74e-6 2.224e-3 45e-6 2.673e-3
MPI_Allreduce 21.873e-3 19.265e-3 17.267e-3 16.373e-3 15.385e-3 14.233e-3 12.625e-3 11.407e-3 10.585e-3
Sum 21.972e-3 19.36e-3 19.066e-3 16.45e-3 16.395e-3 14.307e-3 14.849e-3 11.452e-3 13.258e-3
Mean 10.986e-3 9.67999e-3 9.53301e-3 8.22498e-3 8.19749e-3 7.15349e-3 7.42449e-3 5.726e-3 6.629e-3
StdDev 10.887e-3 9.58499e-3 7.73401e-3 8.14798e-3 7.18749e-3 7.07949e-3 5.20049e-3 5.68101e-3 3.956e-3
22e-3
19.8e-3
17.6e-3
15.4e-3
13.2e-3
11e-3
8.8e-3
6.6e-3
4.4e-3
2.2e-3
0.000 000, 0.149 829: 0.149 829 sec.
All_Processes
Major Function Groups
Tag
Filter

Сравнение результатов трассировки

Окно просмотра **«The Comparison View»** (View Menu -> View -> Compare) позволяет сравнить данные из дву трассировочных файлов или из двух временных интервалов одного трассировочного файла.



Маркировка и фильтрация событий