

С. М. Абрамов, И. М. Загоровский, М. Р. Коваленко,
Г. А. Матвеев, В. А. Роганов

Система визуализации трассы выполнения параллельных программ в среде OpenTS

Аннотация. В статье рассматривается реализация системы визуализации трассы выполнения параллельных программ в среде OpenTS [1]. Подробно рассмотрены основные компоненты системы: бэкенд системы визуализации, являющийся расширением OpenTS и отвечающий за сбор трассы и формирование файла данных трассы в XML-формате и графический фронтенд системы визуализации, предназначенный для анализа и исследования трассы выполнения T-приложения.

Ключевые слова и фразы: T-Система, OpenTS, визуализация, параллельная программа.

Введение

При разработке параллельной T-программы возникает необходимость в понимании того, как происходит ее выполнение: сколько и на каких вычислительных узлах выполняется T-задач, сколько времени выполняются и/или простаивают задачи, какие при этом накладные расходы. Все это представляется важным при отладке и оптимизации программы. С этой целью разработан и используется визуализатор T-программ, состоящий из кода, работающего в среде OpenTS и фиксирующего трассу выполнения параллельной программы, и графического проигрывателя трассы, дающего наглядное представление разработчику о выполнении программы и служащее средством анализа.

Программное обеспечение визуализатора T-программ условно можно разбить на две части:

- средство съема трассы и ее подготовка для последующего отображения в графическом T-визуализаторе;

Работа выполнена при поддержке Программы фундаментальных научных исследований ОИТВС РАН «Оптимизация вычислительных архитектур под конкретные классы задач, информационная безопасность сетевых технологий» и РФФИ (грант РФФИ № 05–07–08005).

- графический T-визуализатор.

1. Алгоритм трассировки

Съем характеристик выполнения программы в среде OpenTS и формирование трассы производится каждую 1/25 секунды. Характеристики снимаются на каждом вычислительном узле, но формирование трассы производится на одном из них, а именно на том, на котором программа была запущена на выполнение — на корневом вычислительном узле. Для передачи характеристик со всех узлов на корневой узел используются MPI-сообщения.

На нулевом узле сообщение принимается и передается функции фиксации трассы.

2. Определение набора характеристик, составляющих трассу

Трасса выполнения T-программы содержит количественные характеристики T-задач в разных состояниях, временные параметры работы планировщика, коммуникационного транспорта, размеры и количество переданных и принятых данных и служебных сообщений.

2.1. Состояния выполнения T-задачи. При выполнении параллельной программы в среде OpenTS T-задача меняет свое состояние от «пренатальной» до «завершенной». Этот процесс является многошаговым. Возможные состояния T-задачи и возможный порядок их изменения приведен на рисунке 1.

Сначала T-задача помещается в очередь пренатальных задач, из которой она может быть экспортирована на другой вычислительный узел, или стать активной. В очередь пренатальных задач также попадают задачи, импортированные с другого вычислительного узла. Выполнение задачи может быть приостановлено в случае отсутствия необходимого готового значения у неготовой переменной. Приостановленные задачи снова переводятся в состояние активных, когда неготовая переменная будет подсчитана до готового значения. Приостановленная задача может быть завершена в случае, если произойдет отказ от результатов ее вычисления.

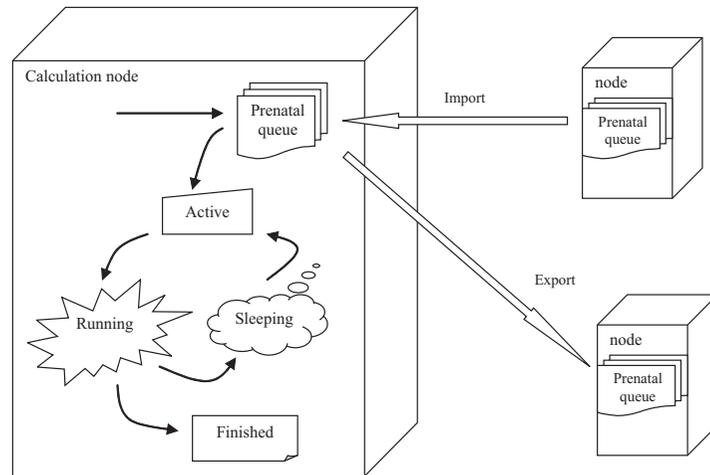


Рис. 1. Состояния Т-задачи

2.2. Типы сообщений. Для передачи данных между вычислительными узлами и управления средой выполнения в OpenTS используются пять типов сообщений. Основными являются сообщения типа «ctrl» и «data».

2.3. Ресурсы вычислительных узлов. В процессе выполнения параллельной программы определяются и вычисляются следующие характеристики вычислительного узла:

- количество свободной памяти, измеряемое в мегабайтах;
- производительность вычислительного узла, измеренная в миллионах операций в секунду (Mflops).

2.4. Временные характеристики. При выполнении программы в среде OpenTS вычисляются следующие временные характеристики:

- суммарное время счета задачи;
- суммарное время простоя задачи;
- суммарное время работы планировщика;
- суммарное время работы MPI-транспорта.

3. Формат представления трассы

Для хранения и передачи трассы выполнения Т-программы графическому визуализатору, разработан формат представления в виде XML.

3.1. XML представление трассы. Заголовок XML-представления трассы содержит информацию о количестве вычислительных узлов, командной строке запуска программы и времени запуска. Например:

```
<trace ranks="4" cmd="fib 10" time="27.04.2006 15:58">
```

Каждую 1/25 секунды формируется временной срез трассы — кадр состояния.

3.2. Кадр состояния. Временной срез трассы содержит номер и время среза, а также детальную информацию по каждому вычислительному узлу. Например:

```
<slice num="3" time="0.143853">
  <rank num="0">
    <resource mbytes="0" mflops="0" tasks="0.003538" idle="0.006406"
      sched="0.003083" mpi="0.002429" />
    <task type="X" count="16" />
    ...
    <detail count="1" to="1" />
    ...
  </task>
  <message sbytes="2536" scount="27" rbytes="1352" rcount="21">
    <extra type="ctrl" sbytes="2536" scount="27" rbytes="1352"
      rcount="21" />
    ...
  </message>
</rank>
...
</slice>
```

Описание характеристик выполнения Т-программы приведено в пункте 4.2.1.

4. Т-визуализатор — графический фронтенд системы визуализации трассы

Программное средство Т-визуализатор является графическим приложением, фронтендом системы визуализации OpenTS.

T-визуализатор предназначен для исследования трассы выполнения T-приложения. Пользователь может использовать T-визуализатор как инструмент для ознакомления с работой T-системы, наблюдая, как происходит распространение и вычисление T-функций, так и для сбора данных, необходимых для отладки и тюнинга T-приложения и T-системы в целом.

T-визуализатор реализован с использованием GTK+ — мультиплатформенного инструментария для создания графических приложений, и является полностью переносимым мультиплатформенным мультисистемным продуктом, работающим как под управлением OS Linux, так и MS Windows, как на платформах i386 (x86, ia32), так и на платформах x86-64 (AMD64, EM64T) [2].

Исходными данными для работы T-визуализатора являются данные трассы выполнения T-приложения. Сбор трассы осуществляется непосредственно в процессе выполнения T-приложения и производится с использованием специальных механизмов OpenTS — расширения сбора трассы и визуализации.

Данные трассы оформляются в виде XML-документа определенной структуры (см. пункт 3.2 данной статьи).

4.1. Режимы работы T-визуализатора. Механизмы обработки исходных данных. Реализованы различные режимы работы T-визуализатора с данными трассы выполнения T-приложения:

- режим T-визуализатора с предварительной загрузкой в память всей трассы выполнения T-приложения;
- потоковая обработка трассы выполнения T-приложения.

4.1.1. *Работа T-визуализатора с предварительно обработанными данными трассы выполнения T-приложения.* В данном режиме производится предварительная обработка данных трассы выполнения T-приложения. То есть, полный разбор XML-файла данных трассы и преобразование данных трассы во внутреннее представление T-визуализатора производится перед началом работы по исследованию трассы.

Преимуществом данного режима является возможность гибкой работы по исследованию трассы, расширенные механизмы управления просмотром трассы: «перемотка» трассы на начало и конец, переход к предыдущему срезу трассы.

Недостатком данного режима является то, что предварительный разбор трассы требует значительных ресурсов при работе с файлами

данных трассы большого объема, и вовсе невозможен, когда объем данных превышает имеющиеся ресурсы установки.

4.1.2. *Потоковая обработка файла данных трассы выполнения T-приложения.* Потоковый режим — это режим, в котором обработка данных трассы (разбор исходного XML-файла с данными) производится постепенно в процессе исследования трассы, по мере необходимости.

Преимуществом данного режима является возможность работы с файлами данных практически любого размера без ограничений.

Недостатком данного режима является то, что отсутствует возможность использования расширенных механизмов управления просмотром трассы, нельзя вернуться к предыдущему срезу.

4.2. Графические компоненты T-визуализатора. Графические компоненты T-визуализатора:

- компонент отображения статистических данных;
- компонент схематического отображения статуса процессов T-приложения;
- компонент управления просмотром трассы T-приложения.

4.2.1. *Компонент отображения статистических данных о состояниях T-приложения.* Компонент предназначен для отображения различного рода информации о состояниях T-приложения в процессе выполнения (см. Рис. 2, нижняя часть).

Информация, отображаемая в статистическом компоненте, разделена на три группы:

- группа «А» — информация по отдельному, выбранному процессу T-приложения;
- группа «Б» — та же информация, что и в группе «А», но суммарная по всем процессам T-приложения;
- группа «В» — содержит неизменяемую информацию, информацию общего характера и информацию о состоянии T-визуализатора.

Группы «А» и «Б» содержат следующую информацию о состояниях T-приложения:

- информацию о состоянии T-функций:
 - количество запущенных T-функций;
 - количество пренатальных T-функций;
 - количество экспортированных T-функций;

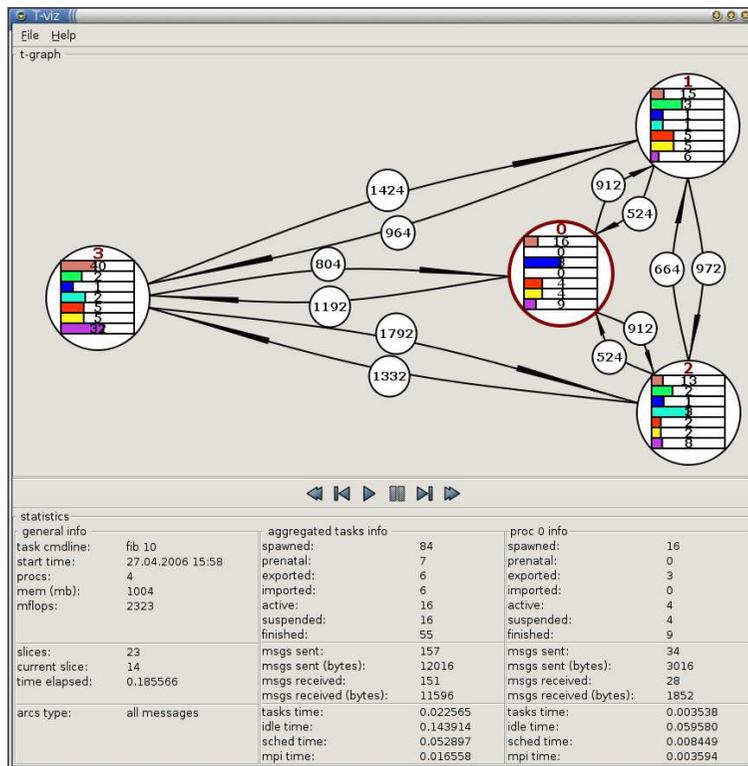


Рис. 2. Графический фронтенд системы визуализации. Компоненты Т-визуализатора. Пример исследования трассы Т-приложения

- количество импортированных Т-функций;
- количество активных (выполняющихся в данный момент) Т-функций;
- количество приостановленных Т-функций;
- количество завершившихся Т-функций.
- информацию о количестве переданных и полученных сообщений;
- информацию о времени работы задачи в той или иной подсистеме OpenTS:

- времени затраченном на выполнение непосредственно Т-функций;
- времени проведенном в планировщике;
- времени потраченном на передачу MPI-сообщений;
- прочее, отображается как время простоя.

Группа «В» содержит информацию:

- об исследуемом Т-приложении, а именно:
 - названии Т-приложения и параметрах запуска;
 - времени старта Т-приложения;
 - количестве процессов;
 - объеме доступной памяти на момент старта Т-приложения;
 - приблизительной оценки вычислительной мощности используемой установки.
- о трассе выполнения Т-приложения:
 - количестве срезов состояния исполнения;
 - текущем срезе;
 - времени, прошедшем с момента старта Т-приложения.
- о типе стрелок компонента схематического отображения статуса процессов Т-приложения.

4.2.2. *Компонент схематического отображения статуса процессов Т-приложения.* Компонент реализован с использованием векторной графической библиотеки Cairo из инструментария GTK+.

Компонент представляет собой набор схематически изображенных процессов рассматриваемого Т-приложения и направленных дуг-стрелок между ними (см. Рис. 2, 3, верхняя часть).

Процесс Т-приложения изображен в виде круга и набора вписанных в него графических элементов отображения статуса, которые отражают информацию по запущенным, пренатальным, экспортированным, импортированным, активным, приостановленным и закончившимся Т-функциям.

Реализована возможность выбора одного из процессов, при этом в статистическом компоненте в подкомпоненте подробной информации по процессу (информация группы «А») будут отображаться данные именно по выбранному процессу.

Реализован механизм перемещения графических подкомпонентов — узлов Т-приложения. Благодаря этому, пользователь может расположить процессы в удобном для исследования виде.

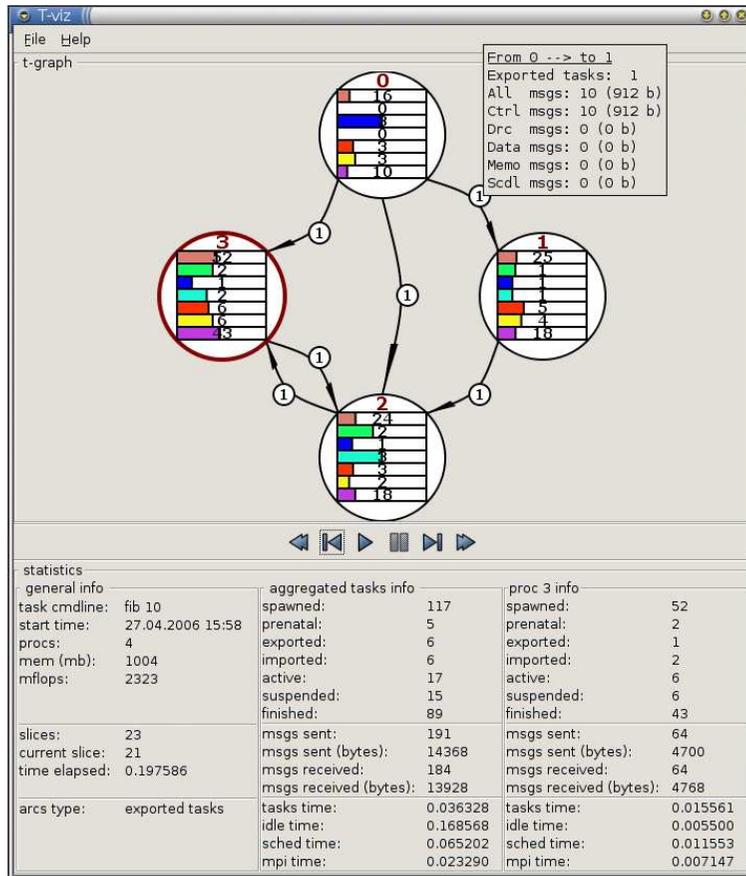


Рис. 3. Иллюстрация миграции Т-задач. Пример информационной всплывающей подсказки

Дуги между процессами отображаются в случае, когда между данными процессами осуществлялся обмен сообщениями или передача Т-функций, в зависимости от выбранного типа дуг. На дугах размещены информационные элементы, отражающие характеристику в зависимости от выбранного типа дуг.

Данный компонент Т-визуализатора имеет механизм всплывающих подсказок (см. Рис. 3, верхняя часть). Всплывающие подсказки

реализованы в виде небольших информационных окон, появляющихся над некоторыми графическими элементами компонента при наведении на них курсора мыши:

- элементами отображения статуса Т-процесса, содержит информацию о том какая часть исследуемой характеристики пришлась на данный процесс;
- информационными элементами дуг, содержит полную информацию о количестве экспортированных Т-функций и информацию по всем типам сообщений между процессами, не зависимо от выбранного типа дуг.

4.2.3. *Компонент управления просмотром трассы Т-приложения.* Компонент управления просмотром позволяет пользователю выбирать интересующий срез трассы для исследования.

Компонент реализован в интуитивно понятном, простом и привычном виде и представляет собой набор графических примитивов — кнопок.

По внешнему виду и способу использования компонент напоминает панель управления некоторым абстрактным проигрывателем (см. Рис. 2, средняя часть).

В наборе присутствуют:

- кнопка «перемотка на начало» — переход на первый срез трассы;
- кнопка «предыдущий кадр» — переход на предыдущий срез;
- кнопка «play» — запустить режим проигрывания (автоматический переход от среза к срезу через заданный промежуток времени);
- кнопка «пауза» — прервать процесс проигрывания;
- кнопка «следующий кадр» — переход на следующий срез;
- кнопка «перемотка на конец» — переход на последний кадр.

Следует отметить, что в режиме потоковой обработки данных трассы доступен не полный набор возможностей управления. Доступные элементы управления в потоковом режиме: включение и выключение режима проигрывания и переход на следующий срез.

4.3. Опции T-визуализатора. Реализован следующий набор опций T-визуализатора:

```
-s, --play_speed=SEC - установить скорость "проигрывания".
                        Например:
                        -s 0.5   - 2 кадра в секунду;
                        -s 0.04  - 25 кадров в секунду.
-f, --file=FILE      - указать файл с xml-данными трассы.
-b, --sbars_style=ID - указать стиль элементов отображения статуса.
                        Возможные варианты: 0 или 1.
-a, --arcs_type=ARCTYPE - указать тип направленных дуг
                        между процессами T-приложения.
                        Возможные варианты:
                        ('e', 'i', 'a', 'c', 'r', 'd', 'm' и 's').
--stream              - использовать потоковый режим обработки данных
                        трассы.
```

Заключение

T-Система реализует автоматическое динамическое распараллеливание программ. Это означает, что многие аспекты организации параллельного счета: распределение задач по узлам кластера, синхронизация процессов, организация пересылки данных, выполняются не программистом, а T-Системой. Во многих случаях T-Системе удается удачно организовать все эти виды работ и получить хороший уровень распараллеливания программ. В тех случаях, когда это не происходит, программист должен понять: как распределялись задачи, почему они не распределялись по узлам, почему было много сообщений и т. д. Для того чтобы исследовать все эти обстоятельства, и была разработана система визуализации, описанная в данной работе. Мы надеемся, что система визуализации будет способствовать повышению удобства использования T-Системы.

Список литературы

- [1] Абрамов С. М., Адамович А. И., Инюхин А. В., Московский А. А., Роганов В. А., Шевчук Ю. В., Шевчук Е. В. *T-система с открытой архитектурой* // Суперкомпьютерные системы и их применение SSA'2004: Труды Международной научной конференции, 26–28 октября 2004 г. Минск, ОИПИ НАН Беларуси. — Минск, 2004, с. 18–22.
- [2] Официальный сайт мультиплатформенного инструментария GTK+: Электронный сетевой ресурс, <http://www.gtk.org>.

ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР МУЛЬТИПРОЦЕССОРНЫХ СИСТЕМ ИПС РАН

S. M. Abramov, M. R. Kovalenko, G. A. Matveev, V. A. Roganov, I. M. Zagorovsky. *Visualization tool of the T-application execution trace.* (in Russian.)

ABSTRACT. In this article we consider the visualization tool of the T-application execution trace.