



Центр
компетенций
и обучения

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

ПРЕДПРИЯТИЕ ГОСКОРПОРАЦИИ «РОСАТОМ»

МССТ 1.0

И НОВЫЕ ВОЗМОЖНОСТИ ПРОВЕРКИ КОРРЕКТНОСТИ ИСПОЛЬЗОВАНИЯ МРІ

А.В. Огородников, М.И. Старов

ООО «Центр Компетенций и Обучения», г. Саров

e-mail: info@compcenter.org

Сложность параллельных комплексов MPI

Несинхронный обмен

Сложные типы данных

Неоднозначность стандарта MPI

Компиляторы и библиотеки MPI не отслеживают ошибки MPI

Вероятность некорректного использования MPI (как следствие, искажение передаваемых данных, зависание программы и т.д.)

Поиск проблем в параллельном программном коде с использованием программ-отладчиков **требует большого количества времени**

МССТ – MPI Correctness Checking Tool

Цель

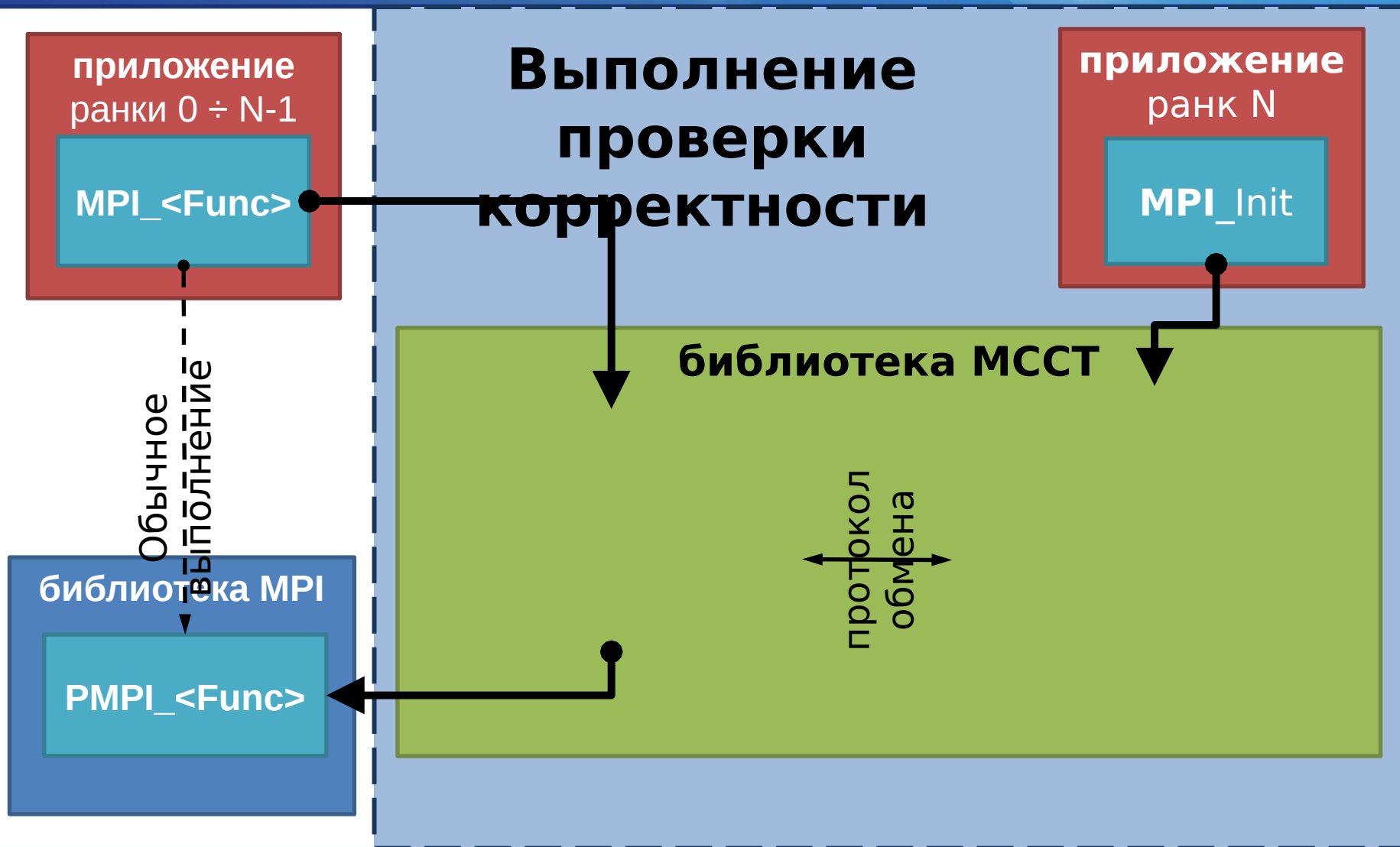
Требования:

- полный охват стандарта MPI-2
- отсутствие ложных срабатываний (false positives)
- ясное для разработчика описание проблем
- указание точного места в программном коде приложения MPI

Имеющиеся конкуренты:

- Intel Trace Analyzer and Collector, Marmot и MUST

Алгоритм проверки корректности



Проверяемые условия

Утилита MCST содержит следующие типы проверок:

лок
аль
ные

1. Соответствие параметров MPI функций их типу и стандарту MPI-2
2. Проверка переносимости исходного кода на другие реализации MPI
3. Проверка целостности используемых данных в вызовах MPI
4. Проверка памяти, используемой пользовательскими ресурсами MPI, на своевременное освобождение
5. Проверка возможного наложения памяти, в т.ч. внутри сложных типов данных MPI

гло
бал
ьн
ые

6. Проверка совпадения типов пересылаемых данных между процессами
7. Сопоставление коллективных операций, вызванных из вовлеченных в данный коммуникатор процессов MPI, по ряду признаков
8. Отслеживание возможных потерь двухточечных сообщений
9. Проверка на условия возникновения взаимоблокировок (deadlocks) между MPI процессами

Сравнение с конкурентами

Список отличий

	МССТ 1.0	Intel TAC 8	MUST 1.1	Marmot 2.4
Запуск без компиляции и линковки	✓	✓	-	-
LD_PRELOAD механизм	✓	✓	✓	-
Отслеживание стека вызовов	✓	✓	✓	-
Проверка целостности передачи данных	✓	✓	-	-
Поддержка MPI-2	Полная	Частичная	Полная	Полная
Интеграция с трассировщиками	В разработке (с XMT)	Intel Trace Collector	-	Vampirtrace
Интеграция с отладчиками	В разработке	✓	-	✓

Особенности MSST 1/2

Можно выделить следующие особенности:

- Отслеживание стека вызовов с использованием библиотеки Stackwalker из состава Dyninst 7.0
- Визуализация взаимоблокировок с использованием утилиты Dot
- Скрипт запуска
`$ checkrun -nr 4 ./a.out`
- Вывод диагностики в STDOUT и HTML-файл

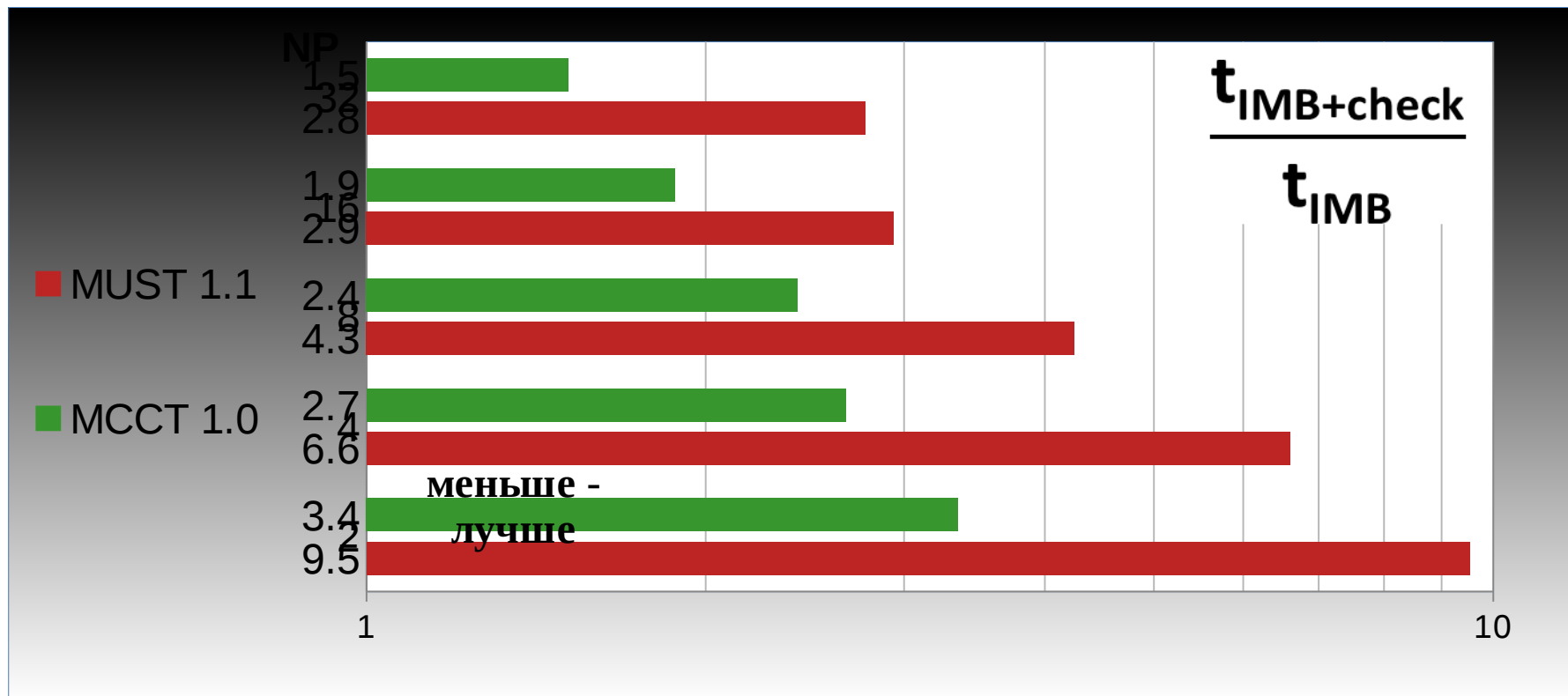
Особенности МССТ 2/2

Отличительные особенности МССТ 1.0:

- Оптимизация базового программного кода MUST 1.1 в наиболее узких местах
- Отсутствие необходимости компиляции и линковки непосредственно перед запуском проверки
- Конфигурирование через окружение
 - ограничение диагностики
`МССТ_CHECK_{TYPE}=on|off`
 - отключение диагностики
`МССТ_{TYPE}_THRESHOLD={number}`
- Проверка целостности пересылаемых данных между процессами
 - Adler-32 для расчета контрольных сумм (Zlib, PHP)
 - сравнение выполняет служебный процесс МССТ

Результат оптимизации

Коэффициент замедления на примере Intel



- Результат проведенных оптимизаций – **ускорение до 3 раз** по сравнению с MUST 1.1

Пример диагностики 1/3

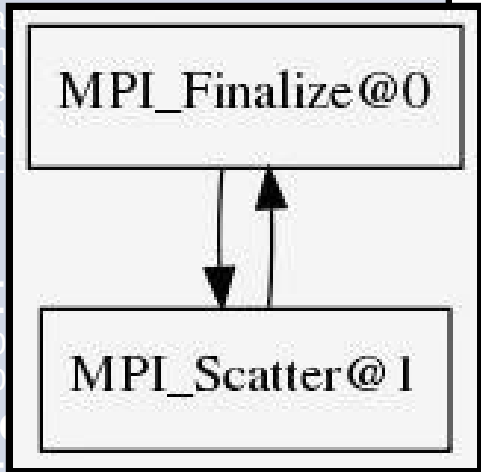
Код примера	Результат проверки
<pre>recvwrongsize.c ... int main (int argc, char **argv) { int rank, buf[10]; ... if(rank == 0) { memset(buf, 1, 10); MPI_Issend(buf, 10, MPI_INT, 1, 100, MPI_COMM_WORLD, &request); buf[0] = 0; MPI_Barrier(MPI_COMM_WORLD); } else if(rank == 1) { MPI_Barrier(MPI_COMM_WORLD); MPI_Recv(buf, 10, MPI_INT, 0, 100, MPI_COMM_WORLD, &status); } ... printf("***END OF TEST***\n"); return 0; }</pre>	<pre>\$ checkrun -n 2 ./recvwrongsize MPI CCT: ERROR: MPI_Recv [rank 1]: MPI CCT: ERROR: Data corruption during transferring from rank 0 to rank 1 MPI CCT: ERROR: MPI_Recv [rank 1] called from: MPI CCT: ERROR: #0 main@recvwrongsize.c:22 MPI CCT: ERROR: #1 __libc_start_main@/lib64/libc-2.12.so:(0x32a2e1ecdd) MPI CCT: ERROR: #2 _start@/home/.../work/checker_examples/recvwrongsize: (0x400809) MPI CCT: INFO: Detected 1 errors and 0 warnings. More details is available in the checker output file. ***END OF TEST***</pre>

Пример диагностики 2/3

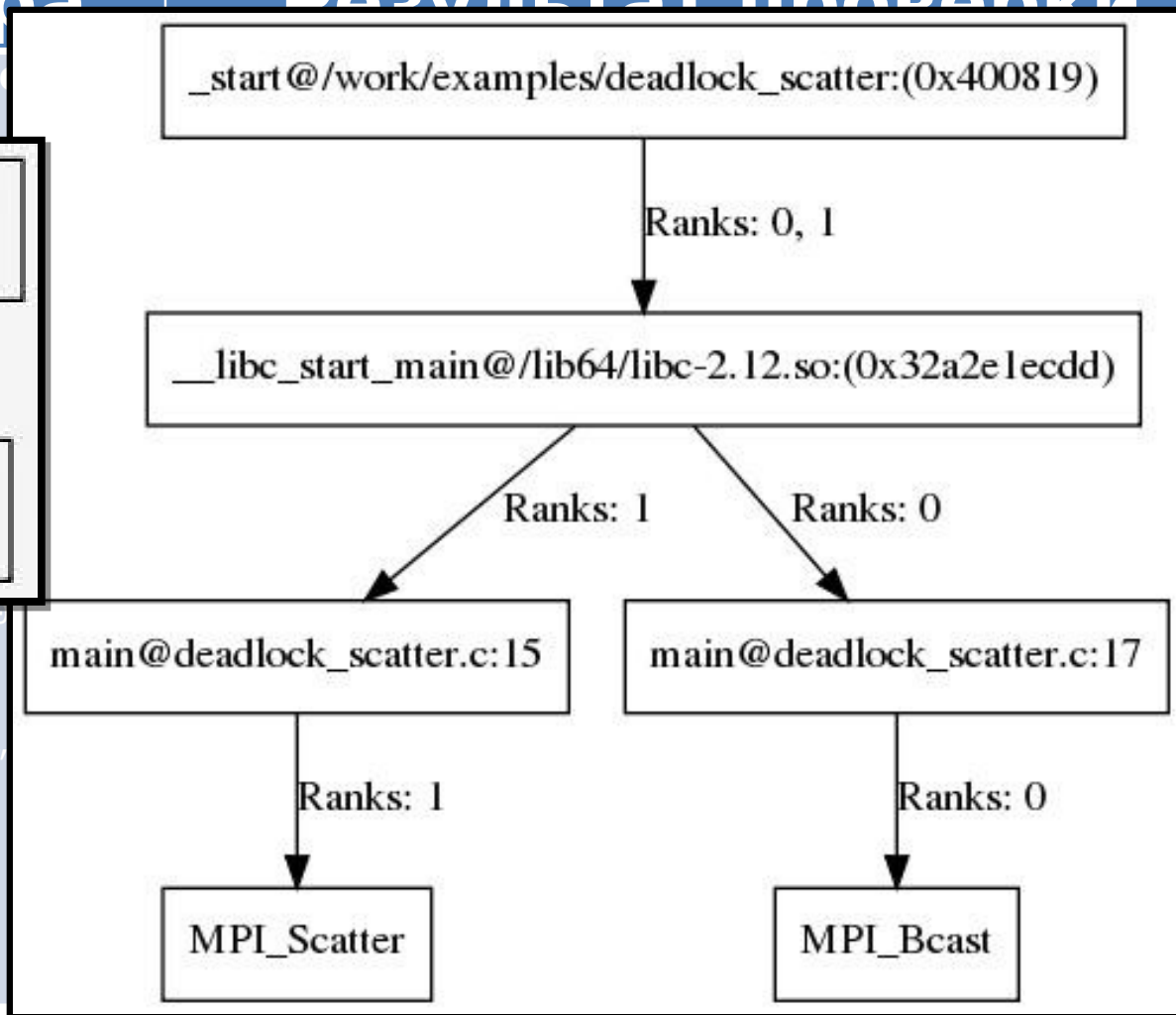
Код примера

deadlock_scatter.c

```
...
int ma
int m
s
recvda
MPI_I
...
if(
MPI
100000
100000
els
MPI_Bcast(&senddata, 10
MPI_INT, 0, MPI_COMM_WORLD);
...
MPI_Finalize();
printf("***END OF TEST***\n"
return 0;
}
```



Результат проверки



Пример диагностики 3/3

Код примера

overlap.c

```
...
int main(int argc, char **argv) {
    int myrank, commsize, data[2];
    MPI_Init( &argc, &argv );
...
    data[0] = myrank + 1;
    data[1] = myrank + 1;

    MPI_Gather(data+1, 1, MPI_INT,
              data, 1, MPI_INT,
              0, MPI_COMM_WORLD );
    if( myrank == 0 )
        printf( "data[0]=%d,data[1]=%d\n",
              data[0], data[1] );
...
    MPI_Finalize( );
    printf("***END OF TEST***\n");
    return 0;
}
```

Результат проверки

```
$ mpicc -g overlap.c -o overlap
$ checkrun -np 2 ./overlap
```

**MPI CCT: ERROR: MPI_Gather [rank 0]:
MPI CCT: ERROR: Overlapping memory
regions of recv and send buffers!**

```
MPI CCT: ERROR: MPI_Gather [rank 0] called from:
MPI CCT: ERROR:   #0 main@overlap.c:16
MPI CCT: ERROR:   #1
__libc_start_main@/lib64/libc-2.12.so:(0x32a2e1ecdd)
MPI CCT: ERROR:   #2
_start@/home/.../work/checker_examples/overlap:
(0x4008a9)
data[0]=1, data[1]=2
```

MPI CCT: INFO: Detected 1 errors and 0 warnings. More details is available in the checker output file.

*** THE END OF TEST ***

Представленный инструмент MCST:

- позволяет сократить время на отладку/оптимизацию сложных параллельных комплексов MPI
- содержит 10 типов различных проверок использования MPI
- позволяет отслеживать стек вызовов и визуализировать взаимоблокировки
- демонстрирует хорошую производительность

Заключение 2/2

В целях развития запланировано следующее

- переход на новые базовые версии MUST 1.2+
- поддержка стандарта MPI-3
- интеграция MCST с компонентой профилирования и трассировки ХМТ в составе ПК S-MPI
- интеграция с отладчиками GDB, Alinea DDT, TotalView

**Спасибо за внимание.
Вопросы?**



ООО «Центр компетенций и обучения»

info@compcenter.org

(83130) 6-70-67