

# 分布式系统作业

## 第 6 次作业

姓名：唐晨轩

班级：人工智能与大数据

学号：19335182

### 一、 问题描述

#### Homework-6

一、找出给出的7个MPI程序中的bug，并将程序修改正确，比较修改前后的程序运行结果。MPI程序见附件。

二、利用MPI程序实现  $n$  个整形数的排序过程，排序算法不限，MPI的进程不限，可以使用单机运行多个MPI进程。提示：在排序的过程中需要MPI集合通信如 `MPI_Allgather` 等。

### 二、 解决方案

1.

程序 1:

`rank = 0` 和 `rank = 1` 的两个进程，进程中的 `send` 和 `recv` 函数的 `tag` 不一致，所以

二者不能实现消息传递。

修改前:

```
hadoop@ubuntu: ~/hw6/home work 6/home work 6$ mpirun -np
4 ./mpi_bug1
Task 0 starting...
Numtasks=4. Only 2 needed. Ignoring extra...
Sent to task 1...
Task 3 starting...
Task 1 starting...
Task 2 starting...
█
```

修改后:

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug1
Task 1 starting...
Task 3 starting...
Task 0 starting...
Numtasks=4. Only 2 needed. Ignoring extra...
Sent to task 1...
Sent to task 0...
Received from task 1...
Task 1: Received 1 char(s) from task 0 with tag 1
Task 0: Received 1 char(s) from task 1 with tag 2
Task 2 starting...
```

程序 2:

MPI\_Isend 和 MPI\_Irecv 的 buf\_byte 不同, 一个是 MPI\_INT, 另一个是 MPI\_FLOAT.

修改前:

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug2
Numtasks=4. Only 2 needed. Ignoring extra...
Task 0 sent = 0
Task 0 sent = 10
Task 0 sent = 20
Task 0 sent = 30
Task 0 sent = 40
Task 0 sent = 50
Task 0 sent = 60
Task 0 sent = 70
Task 0 sent = 80
Task 0 sent = 90
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
```

修改后:

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug2
Numtasks=4. Only 2 needed. Ignoring extra...
Task 0 sent = 0
Task 0 sent = 10
Task 0 sent = 20
Task 0 sent = 30
Task 0 sent = 40
Task 0 sent = 50
Task 0 sent = 60
Task 0 sent = 70
Task 0 sent = 80
Task 0 sent = 90
Task 1 received = 0
Task 1 received = 10
Task 1 received = 20
Task 1 received = 30
Task 1 received = 40
Task 1 received = 50
Task 1 received = 60
Task 1 received = 70
Task 1 received = 80
Task 1 received = 90
hadoop@ubuntu: ~/hw6/home work6/home work6$
```

程序 3:

缺少函数 MPI\_INIT 和 MPI\_Finalize。

修改前：

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug3
Attempting to use an MPI routine before initializing MPI CH
Attempting to use an MPI routine before initializing MPI CH
Attempting to use an MPI routine before initializing MPI CH
Attempting to use an MPI routine before initializing MPI CH
hadoop@ubuntu: ~/hw6/home work6/home work6$
```

修改后：

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug3
MPI task 2 has started...
MPI task 0 has started...
MPI task 1 has started...
MPI task 3 has started...
Initialized array sum = 1.335708e+14
Sent 4000000 elements to task 1 offset= 4000000
Task 1 mysum = 4.884048e+13
Sent 4000000 elements to task 2 offset= 8000000
Sent 4000000 elements to task 3 offset= 12000000
Task 2 mysum = 7.983003e+13
Task 0 mysum = 1.598859e+13
Task 3 mysum = 1.161867e+14
Sample results:
 0.000000e+00  2.000000e+00  4.000000e+00  6.000000e+00  8.000000e+00
 8.000000e+06  8.000002e+06  8.000004e+06  8.000006e+06  8.000008e+06
 1.600000e+07  1.600000e+07  1.600000e+07  1.600001e+07  1.600001e+07
 2.400000e+07  2.400000e+07  2.400000e+07  2.400001e+07  2.400001e+07
*** Final sum= 2.608458e+14 ***
hadoop@ubuntu: ~/hw6/home work6/home work6$
```

程序 4：

所有进程结束后归约操作没有用 MPI\_Reduce，所以 sum 的结果不对。

修改前：

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug4
MPI task 0 has started...
MPI task 3 has started...
MPI task 1 has started...
MPI task 2 has started...
Initialized array sum = 1.335708e+14
Sent 4000000 elements to task 1 offset= 4000000
Sent 4000000 elements to task 2 offset= 8000000
Task 1 mysum = 4.884048e+13
Task 2 mysum = 7.983003e+13
Sent 4000000 elements to task 3 offset= 12000000
Task 3 mysum = 1.161867e+14
Task 0 mysum = 1.598859e+13
Sample results:
 0.000000e+00  2.000000e+00  4.000000e+00  6.000000e+00  8.000000e+00
 8.000000e+06  8.000002e+06  8.000004e+06  8.000006e+06  8.000008e+06
 1.600000e+07  1.600000e+07  1.600000e+07  1.600001e+07  1.600001e+07
 2.400000e+07  2.400000e+07  2.400000e+07  2.400001e+07  2.400001e+07
*** Final sum= 1.335708e+14 ***
hadoop@ubuntu: ~/hw6/home work6/home work6$
```

修改后：

```

hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug4
MPI task 3 has started...
MPI task 0 has started...
MPI task 1 has started...
MPI task 2 has started...
Initialized array sum = 1.335708e+14
Sent 4000000 elements to task 1 offset= 4000000
Task 1 mysum = 4.884048e+13
Sent 4000000 elements to task 2 offset= 8000000
Sent 4000000 elements to task 3 offset= 12000000
Task 2 mysum = 7.983003e+13
Task 0 mysum = 1.598859e+13
Task 3 mysum = 1.161867e+14
Sample results:
 0.000000e+00  2.000000e+00  4.000000e+00  6.000000e+00  8.000000e+00
 8.000000e+06  8.000002e+06  8.000004e+06  8.000006e+06  8.000008e+06
 1.600000e+07  1.600000e+07  1.600000e+07  1.600001e+07  1.600001e+07
 2.400000e+07  2.400000e+07  2.400000e+07  2.400001e+07  2.400001e+07
*** Final sum= 2.608458e+14 ***
hadoop@ubuntu: ~/hw6/home work6/home work6$

```

程序 5:

由于接收消息的进程 rank1 需要花费比发送消息的进程 rank0 更多的时间，这里用的发送和接受函数均为阻塞式，所以会导致发送消息的 rank0 进入阻塞状态从而造成不合理的时间安排。

修改前:

```
Count= 330   Time= 0.136219 sec.
Count= 340   Time= 0.142980 sec.
Count= 350   Time= 0.172364 sec.
Count= 360   Time= 0.165734 sec.
Count= 370   Time= 0.151263 sec.
Count= 380   Time= 0.123997 sec.
Count= 390   Time= 0.127222 sec.
Count= 400   Time= 0.139225 sec.
Count= 410   Time= 0.147403 sec.
Count= 420   Time= 0.130824 sec.
Count= 430   Time= 0.122399 sec.
Count= 440   Time= 0.120187 sec.
Count= 450   Time= 0.136197 sec.
Count= 460   Time= 0.132530 sec.
Count= 470   Time= 0.143589 sec.
Count= 480   Time= 0.120357 sec.
Count= 490   Time= 0.146483 sec.
Count= 500   Time= 0.135066 sec.
Count= 510   Time= 0.138973 sec.
Count= 520   Time= 0.116134 sec.
Count= 530   Time= 0.126199 sec.
Count= 540   Time= 0.140214 sec.
Count= 550   Time= 0.147403 sec.
Count= 560   Time= 0.155153 sec.
Count= 570   Time= 0.148454 sec.
Count= 580   Time= 0.158642 sec.
Count= 590   Time= 0.166599 sec.
Count= 600   Time= 0.159283 sec.
Count= 610   Time= 0.147197 sec.
Count= 620   Time= 0.162883 sec.
Count= 630   Time= 0.126374 sec.
Count= 640   Time= 0.151846 sec.
Count= 650   Time= 0.197588 sec.
Count= 660   Time= 0.152905 sec.
Count= 670   Time= 0.160079 sec.
Count= 680   Time= 0.163437 sec.
```

修改后:

程序 6:

修改前只有 Isend/Irecv 的组合成功完成收发消息。因为 MPI\_Waitall 函数有错误, 所以 MPI\_Irecv 的 buf 不能真正接收到数据。

修改前:

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug6
Starting isend/irecv send/irecv test...
Task 0 starting...
Task 1 starting...
Task 2 starting...
Task 3 starting...
Task 0 has done 100 isends/irecvs
Task 0 has done 200 isends/irecvs
Task 0 has done 300 isends/irecvs
Task 0 has done 400 isends/irecvs
Task 0 has done 500 isends/irecvs
Task 0 has done 600 isends/irecvs
Task 0 has done 700 isends/irecvs
Task 0 has done 800 isends/irecvs
Task 0 has done 900 isends/irecvs
Task 0 has done 1000 isends/irecvs
Task 1 has done 100 isends/irecvs
Task 1 has done 200 isends/irecvs
Task 1 has done 300 isends/irecvs
Task 1 has done 400 isends/irecvs
Task 1 has done 500 isends/irecvs
Task 1 has done 600 isends/irecvs
Task 1 has done 700 isends/irecvs
Task 1 has done 800 isends/irecvs
Task 1 has done 900 isends/irecvs
Task 1 has done 1000 isends/irecvs
Fatal error in PMPI_Waitall: Request pending due to failure, error stack:
PMPI_Waitall(356): MPI_Waitall(count=2000, req_array=0x7fff24b14aa0, status_array=0x7fff24b169e0) failed
PMPI_Waitall(332): The supplied request in array element 0 was invalid (kind=0)
```

修改后:

程序 7:

MPI\_Bcast 广播函数第二个参数有误, 应该为 1

修改前:

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug7
Task 1 on ubuntu starting...
Task 0 on ubuntu starting...
Root: Number of MPI tasks is: 4
Task 3 on ubuntu starting...
Task 2 on ubuntu starting...
█
```

修改后:

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./mpi_bug7
Task 2 on ubuntu starting...
Task 3 on ubuntu starting...
Task 0 on ubuntu starting...
Root: Number of MPI tasks is: 4
Task 1 on ubuntu starting...
hadoop@ubuntu: ~/hw6/home work6/home work6$ █
```

2.

排序算法选择了奇偶排序算法。

当仅仅使用 1 个进程的时候进入串行代码块, 否则进入并行代码块:

串行时长:

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpic++ oe_sort.cpp -o oe_sort
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 1 ./oe_sort
串行时间 0.224789
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 1 ./oe_sort
串行时间 0.223767
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 1 ./oe_sort
串行时间 0.228675
hadoop@ubuntu: ~/hw6/home work6/home work6$
```

4 个进程并行时长：

```
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.217184
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.224498
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.174089
hadoop@ubuntu: ~/hw6/home work6/home work6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.200328
hadoop@ubuntu: ~/hw6/home work6/home work6$
```

排序数字个数为 9900 个，结果表现串行并行输出差异不大，我认为原因在于：

一方面输出本身比较耗时；另一方面多进程抢占输出设备资源，互相等待；

**代码在压缩包中附带有。**