### Possible deadlock with tmpi and pin=auto

**08/08/2016 11:56 PM - Roland Schulz**

**Status:** Closed  
**Priority:** High  
**Assignee:**  
**Category:** mdrun  
**Target version:** 2016.1  
**Affected version:** 2016  
**Difficulty:** uncategorized

**Description**

It is possible that `bAllSet` isn't the same for all tmpi ranks at `threadaffinity.cpp:525`. This in turn than causes a deadlock because setting the affinity requires global communication. This can't happen with lib-MPI because in that case `MPI_Allreduce` is used. I suspect that it isn't the same for all ranks because the affinity gets changed by some threads (not sure whether by OpenMP/MPI or by GROMACS) while others test.

Not sure what the best solution is because tmpi isn't initialized yet at that spot. Thus one cannot simply do a `MPI_Barrier` or `MPI_Allreduce`. 

This is with ICC 17beta1 on KNL but should be possible to reproduce on other compiler/CPUs.

**Associated revisions**

Revision 82216120 - 08/10/2016 07:50 AM - Berk Hess  
Fix deadlock with thread-MPI  

With thread-MPI mdrun could deadlock while pinning threads.

Fixes #2025.  
Change-Id: Ib42e9625134531b1e2f910b11339aa0f78b80624

**History**

#### #1 - 08/09/2016 02:21 PM - Berk Hess

How can `bAllSet` be different with tmpi? All tmpi threads will be running on the same hardware and I don't see any input that depends on the tmpi thread index.

#### #2 - 08/09/2016 05:59 PM - Teemu Murtola

Different ranks call the function (and so also `sched_getaffinity()`) at different times, so if some library call has changed the affinity (most likely between the first rank and subsequent ranks), then this could possibly happen.

#### #3 - 08/09/2016 06:01 PM - Roland Schulz

Exactly. If I print `bAllSet` I see that different ranks have different values. Because it depends on timing it doesn't always happen. But most of the time.

#### #4 - 08/09/2016 09:29 PM - Berk Hess

Shouldn't this check be done only on a single thread with thread-MPI?  
If so, we can simply move the return at 487 out of the `!bAfterOpenmpInit` conditional.

#### #5 - 08/09/2016 10:19 PM - Berk Hess

I think I see what the issue is now. In the after-OpenMP-init call, one thread is probably changing the affinities while others are detecting. The solution is to simply do the reduction also with thread-MPI after openMP init. I will upload a fix.

#### #6 - 08/09/2016 10:37 PM - Gerrit Code Review Bot

Gerrit received a related patchset '1' for Issue #2025.  
Uploader: Berk Hess ([hess@kth.se](mailto:hess@kth.se))  
Change-Id: Ib42e9625134531b1e2f910b11339aa0f78b80624  
Gerrit URL: [https://gerrit.gromacs.org/6107](https://gerrit.gromacs.org/6107)
#7 - 08/09/2016 10:38 PM - Berk Hess
- Category set to mdrun
- Status changed from New to Fix uploaded
- Priority changed from Normal to High
- Target version set to 2016.1

#8 - 08/12/2016 10:08 AM - Berk Hess
- Status changed from Fix uploaded to Resolved

Applied in changeset 82216120d0f7a58289291a68256845816558f115.

#9 - 10/19/2016 02:07 PM - Szilárd Páll
- Status changed from Resolved to Closed