

Re: [ofa-general] Re: [PATCH RFC] RDMA/CMA: Allocate PS\_TCPportsfrom the host TCP port space.

## Re: [ofa-general] Re: [PATCH RFC] RDMA/CMA: Allocate PS\_TCPportsfrom the host TCP port space.

---

*Source:* <http://linux.derkeiler.com/Mailing-Lists/Kernel/2007-08/msg08870.html>

---

- *From:* Patrick Geoffray <patrick@xxxxxxxx>
  - *Date:* Mon, 20 Aug 2007 16:33:56 -0400
- 

Felix Marti wrote:

Yes, the app will take the cache hits when accessing the data. However, the fact remains that if there is a copy in the receive path, you require an additional 3x memory BW (which is very significant at these high rates and most likely the bottleneck for most current systems)... and somebody always has to take the cache miss be it the copy\_to\_user or the app.

The cache miss is going to cost you half the memory bandwidth of a full copy. If the data is already in cache, then the copy is cheaper.

However, removing the copy removes the kernel from the picture on the receive side, so you lose demultiplexing, asynchronism, security, accounting, flow-control, swapping, etc. If it's ok with you to not use the kernel stack, then why expect to fit in the existing infrastructure anyway ?

Yes, RDMA support is there... but we could make it better and easier to

What do you need from the kernel for RDMA support beyond HW drivers ? A fast way to pin and translate user memory (ie registration). That is pretty much the sandbox that David referred to.

Eventually, it would be useful to be able to track the VM space to implement a registration cache instead of using ugly hacks in user-space to hijack malloc, but this is completely independent from the net stack.

use. We have a problem today with port sharing and there was a proposal

The port spaces are either totally separate and there is no issue, or completely identical and you should then run your connection manager in user-space or fix your middlewares.

Re: [ofa-general] Re: [PATCH RFC] RDMA/CMA: Allocate PS\_TCPportsfrom the host TCP port space.

Re: [ofa-general] Re: [PATCH RFC] RDMA/CMA: Allocate PS\_TCPportsfrom the host TCP port space.

and not for technical reasons. I believe this email threads shows in detail how RDMA (a network technology) is treated as bastard child by the network folks, well at least by one of them.

I don't think it's fair. This thread actually show how pushy some RDMA folks are about not acknowledging that the current infrastructure is here for a reason, and about mistaking zero-copy and RDMA.

This is a similar argument than the TOE discussion, and it was definitively a good decision to not mess up the Linux stack with TOEs.

Patrick

—

To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@xxxxxxxxxxxxxxxxx

More majordomo info at <http://vger.kernel.org/majordomo-info.html>

Please read the FAQ at <http://www.tux.org/lkml/>

Re: [ofa-general] Re: [PATCH RFC] RDMA/CMA: Allocate PS\_TCPportsfrom the host TCP port space.