

Re: [PATCH][RFC][0/4] InfiniBand userspace verbs implementation

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2005-04/6149.html>

From: Andy Isaacson (adi_at_hexapodia.org)

Date: 04/25/05

Date: Mon, 25 Apr 2005 12:11:11 -0700

To: Andrew Morton <akpm@osdl.org>

On Sat, Apr 23, 2005 at 07:44:21PM -0700, Andrew Morton wrote:

> Timur Tabi <timur.tabi@ammasso.com> wrote:
> > As I said, the testcase only works with our hardware, and it's also
> > very large. It's one small test that's part of a huge test suite.
> > It takes a couple hours just to install the damn thing.
> >
> > We want to produce a simpler test case that demonstrates the problem in an
> > easy-to-understand manner, but we don't have time to do that now.
>
> If your theory is correct then it should be able to demonstrate this
> problem without any special hardware at all: pin some user memory, then
> generate memory pressure then check the contents of those pinned pages.
>
> But if, for the DMA transfer, you're using the array of page*'s which were
> originally obtained from `get_user_pages()` then it's rather hard to see how
> the kernel could alter the page's contents.
>
> Then again, if `mlock()` fixes it then something's up. Very odd.

Andrew,

Libor Michalek posted a much more reasonable (to my limited understanding) bug description in <20050412180447.E6958@topspin.com>.

(And I'd love to provide a URL, but damned if I can figure out how to find that message on gmane. Clue-bat applications gladly accepted.)

Libor Michalek wrote:

```
# The driver did use get_user_pages() to elevated the refcount on all the
# pages it was going to use for IO, as well as call set_page_dirty() since
# the pages were going to have data written to them from the device.
#
# The problem we were seeing is that the minor fault by the app resulted
# in a new physical page getting mapped for the application. The page that
# had the elevated refcount was still waiting for the data to be written
```

Linux-Kernel: Re: [PATCH][RFC][0/4] InfiniBand userspace verbs implementation

```
# to by the driver at the time that the app accessed the page causing the  
# minor fault. Obviously since the app had a new mapping the data written  
# by the driver was lost.  
#  
# It looks like code was added to try_to_unmap_one() to address this, so  
# hopefully it's no longer an issue...
```

Which makes me think that Timur's bug is just an
insufficiently-understood version of Libor's.

-andy

-

To unsubscribe from this list: send the line "unsubscribe linux-kernel" in
the body of a message to majordomo@vger.kernel.org
More majordomo info at <http://vger.kernel.org/majordomo-info.html>
Please read the FAQ at <http://www.tux.org/lkml/>