

Re: [PATCH] get_nodes mask miscalculation

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2004-08/2759.html>

From: Brent Casavant (bcasavan_at_sgi.com)

Date: 08/10/04

Date: Tue, 10 Aug 2004 09:54:08 -0500
To: Andi Kleen <ak@muc.de>

On Tue, 10 Aug 2004, Andi Kleen wrote:

- > *The idea behind this is was to make it behave like select.*
- > *And select passes highest valid value + 1.*

Not that my opinion is worth much, but this seems a very peculiar interface to mimic.

- > *In this case*
- > *valid value is not the highest bit number, but the length*
- > *of the bitmap.*
- >
- > *For some reason nobody except me seems to get it though,*
- > *probably because the description in the manpages is a bit confusing:*
- >
- > *get_mempolicy(2):*
- >
- > *"maxnode is the maximum bit number plus one that can be stored into*
- > *nodemask."*

Actually, I worked from the following description:

```
mbind(3)
...
nodemask is a bit field of nodes that contains up to
maxnode bits.
...
```

This very clearly indicates that I should pass the number of bits in the nodemask field, not the number of bits plus one.

Granted, the man page makes it clear that I should use libnuma, but in this case I was working on implementing a new MPOL_ROUNDROBIN, and wanted to go directly to the system call for testing purposes.

And not to make things even more confusing, but the way things are designed now, the value I need to pass to mbind() is `numa_max_node()+2`.

Linux-Kernel: Re: [PATCH] get_nodes mask miscalculation

Very confusing.

- > *Problem is that you'll break all existing libnuma binaries*
- > *which pass NUMA_MAX_NODES + 1. In your scheme the kernel*
- > *will access one bit beyond the bitmap that got passed,*
- > *and depending on its random value you may get a failure or not.*

Well, again, not that my opinion carries much weight (nor really should it), but this whole off-by-one (or two) interface seems unnecessarily confusing. Is there no way we can get this fixed? The temporary pain of breaking the relatively new libnuma seems worth the price of getting this forever cleaned up.

Brent

--

Brent Casavant