Subject: Re: Re: [RFC][PATCH 2/7] RSS controller core
Posted by Paul Menage on Sun, 18 Mar 2007 22:44:24 GMT
View Forum Message <> Reply to Message

On 3/13/07, Dave Hansen <hansendc@us.ibm.com> wrote:
> How do we determine what is shared, and goes into the shared zones?
> Once we've allocated a page, it's too late because we already picked.
> Do we just assume all page cache is shared? Base it on filesystem,
> mount, ...? Mount seems the most logical to me, that a sysadmin would
> have to set up a container's fs, anyway, and will likely be doing
> special things to shared data, anyway (r/o bind mounts :).

I played with an approach where you can bind a dentry to a set of memory zones, and any children of that dentry would inherit the mempolicy; I was envisaging that most data wouldn't be shared between different containers/jobs, and that userspace would set up "shared" zones for big shared regions such as /lib, /usr, /bin, and for specially-known cases of sharing.

> If we really do bind a set of processes strongly to a set of memory on a
> set of nodes, then those really do become its home NUMA nodes. If the
> CPUs there get overloaded, running it elsewhere will continue to grab
> pages from the home. Would this basically keep us from ever being able
> to move tasks around a NUMA system?

move_pages() will let you shuffle tasks from one node to another without too much intrusion.

Paul

____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers