Subject: Re: [ckrm-tech] [PATCH 1/1] Fix a panic while mouting containers on powerpc and some other small cle
Posted by Balbir Singh on Mon, 15 Jan 2007 09:51:28 GMT
View Forum Message <> Reply to Message

Paul Menage wrote:
> On 1/15/07, Balbir Singh <balbir@in.ibm.com> wrote:
>> In sched.c, account_user_time() can be called with the task p set to rq->idle.
>> Since idle tasks do not belong to any container, this was causing a panic in
>> task_ca() in cpu_acct.c.
>
> How come that didn't cause a problem on x86_64? If this is an
> inconsistency between architectures then perhaps it ought to be
> cleaned up.
>

That is because account_system/user_time() is also called from
account_process_vtime() which is called from __switch_to in
power pc. vtime is for virtual time accounting. Enabled by
CONFIG_VIRT_CPU_ACCOUNTING.

> Additionally, I think that we should make the idle tasks members of
> the root container(s), to remove this special case. (I'm a bit
> surprised that they're not already - I thought that the early
> container initialization was early enough that the idle tasks hadn't
> yet been forked. Is that different on PowerPC?
>

idle threads are associated only with the runqueue and not visible
by the do_each_thread()/while_each_thread() loop. They are not added
to the tasklist (please see init_idle() in kernel/sched.c).

>> Multiplying the time by 1000 is not correct in cpuusage_read(). The code
>> has been converted to use the correct cputime API.
>
> Thanks.
>
>> Add mount/umount callbacks.
>
> I'm not sure I like the mount/unmount callbacks. What exactly are you
> trying to gain from them? My intention was that the
>
> cont->subsys[i]->container = cont;
>
> line in container_get_sb() was doing essentially this - i.e. the
> container_subsys_state for the root container in a subsystem is
> already kept up to date by the container system, and the subsystem can
> rely on the "container" field in the container_subsys_state.

>

While writing/extending the cpuacct container, I found it useful to
know if the container resource group we are controlling is really mounted.
Controllers can try and avoid doing work when not mounted and start
when the subsystem is mounted. Also, without these callbacks, one has no
definite way of checking if the top_container is dummy or for real.

> Thanks,
>
> Paul

--

Balbir Singh,
Linux Technology Center,
IBM Software Labs