Subject: Re: [PATCH] Mark timer_stats as incompatible with multiple pid namespaces
Posted by ebiederm on Thu, 13 Dec 2007 11:55:22 GMT
View Forum Message <> Reply to Message

Ingo Molnar <mingo@elte.hu> writes:

> * Eric W. Biederman <ebiederm@xmission.com> wrote:
>
>> /proc/timer_stats currently reports the user of a timer by pid, which
>> is a reasonable approach. However if you are not in the initial pid
>> namespace the pid that is reported is nonsense.
>>
>> Therefore until we can make timer_stats pid namespace safe just
>> disable it in the build if pid namespace support is selected so we at
>> least know we have a conflict.
>
> What the heck??? Please solve this properly instead of hiding it.
> /proc/timer_stats is damn useful and it's a must-have for powertop to
> work.

Hmm. Perhaps the dependency conflict should go in the other direction
then.

My goal is to document the issue while a proper fix is being written.
I have known about this for all of about 1 day now. It was added since
last time I went through the kernel and made a thorough sweep of pid users.

What the proper fix is isn't even obvious at this point.
Possibly it is making /proc/timer_stats disappear in child pid namespaces.
  Which we don't currently have the infrastructure fore.
Possibly it is reworking the stats collection so we store a struct pid *
instead of a pid_t value. So we would know if the reader of the value
can even see processes you have collected stats for.

It is going to take a bit to digest what is going on and solve this properly.

In the same vein do we actively have interesting user space programs
using /proc/sched_debug? It is the same class of problem. Yet
another interface talking to user space with pids.

Eric

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