
Subject: Re: [PATCH v4 0/4] per cgroup cpu statistics
Posted by [Glauber Costa](#) on Thu, 14 Jun 2012 12:18:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 06/05/2012 06:49 PM, Glauber Costa wrote:

> Hello,
>
> This is the 4th version of this per-cgroup statistics. My aim with that is
> to provide userspace with the necessary tools to export a view of /proc/stat
> inside a container, so tools like top would work.
>
> I believe this version merges all the comments from Peter and Paul. Let me know
> if I forgot any of them.
>
> I now reuse put_prev_task to account for nr_switches, so no extra walks are added.
> For that, I had to add another parameter to the function, but I hope this is
> acceptable. Please note that this parameter would still be needed even if
> Peter's patch that merges put_prev_task behavior inside pick_next_task.
>
> This is because the previous class may live in a different class and we may
> have to call it anyway. In which case a hint like this would be needed.
> I believe this to be orthogonal to the work you are doing.
>
> Peter: the patch that adds exec_clock to rt was removed. That was my bad, it
> was intended to be part of the series that unifies cpu and cpuacct for the
> comounted case. I have them together in my tree, and I made the cut in the wrong
> place. It is not needed here.
>
> Let me know what you think of this.
>
> v4:
> * read_seq_string used instead of read_map. This is because the
> buffer can get very big, and that's easier to fix by using seq_string
> * idle time no longer exported. It can be derived from userspace easily
> * "steal" changed to "wait", since steal is more our interpretation of it
> * nr_switches now being accounted as we walk the tree in put_prev_task, so no
> new hierarchy walks are being inserted.
> v3:
> * completely reworked nr_switches gathering
> * separated per-se sleep_start to be more clear about it
>
>
Any comments in this incarnation of the series?
