Subject: How to draw values for /proc/stat
Posted by Glauber Costa on Mon, 05 Dec 2011 09:32:33 GMT
View Forum Message <> Reply to Message

Hi,

Specially Peter and Paul, but all the others:

As you can see in https://lkml.org/lkml/2011/12/4/178, and in my answer to that, there is a question - one I've asked before but without that much of an audience - of whether /proc files read from process living on cgroups should display global or per-cgroup resources.

In the past, I was arguing for a knob to control that, but I recently started to believe that a knob here will only overcomplicate matters: if you live in a cgroup, you should display only the resources you can possibly use. Global is for whoever is in the main cgroup.

Now, it comes two questions:
1) Do you agree with that, for files like /proc/stat ? I think the most important part is to be consistent inside the system, regardless of what is done

2) Will cpuacct stay? I think if it does, that becomes almost mandatory (at least the bind mount idea is pretty much over here), because drawing value for /proc/stat becomes quite complex.
The cpuacct cgroup can provide user, sys, etc values. But we also have:

* nr_context_switches,
* jiffies since boot,
* total_forks,
* nr_running,
* nr_iowait,

Now I doubt any of us want to see /proc/stat extended to accommodate things like nr_context_switches, or even worse, nr_running. The way I see it, there are two options here:

a) moving everything to cpu cgroup so we keep all values being drawn from the same place
b) Collect that info from multiple places in a transparent way. ctx, nr_running and nr_iowait will probably come from cpu. jiffies can come from wherever, and maybe we can even draw total_forks from Frederic's and avoid counting it twice.