Subject: Re: How to draw values for /proc/stat
Posted by Glauber Costa on Tue, 06 Dec 2011 00:17:06 GMT

View Forum Message <> Reply to Message

On 12/05/2011 10:05 PM, KAMEZAWA Hiroyuki wrote:
> On Mon, 5 Dec 2011 07:32:33 -0200
> Glauber Costa<glommer@parallels.com> wrote:
>
>> 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.
>>
>
> Hm. I have a suggestion and a concern.
>
> (A suggestion)
>   How about having a mount option for procfs ?
>   For example,
> mount -t proc .... -o cgroup_virtualized
>   Then, /proc/stat etc shows per-cgroup information.
>
> (A concern)
>   /proc/stat will be a mixture of virtualized values and not-virtualized values.
>   1. Don't users need to know whether each value is virtualized or not ?
>   2. Can we have a way to show "this value is virtualized!" annotation ?

A mount options works for me.
However, "work" doesn't mean it is really necessary, and that's the real
question: is there a real use case for someone resource-constrained to
see system-wide values ?

What is exactly the expectation here? If you want to see whichever
resources you're entitled to, just showing the cgroups version on /proc
would be simpler, with less confusion, and do the right thing.

As for your concerns:
1) As said above, my point of view is that no, they do not. You only see
what you can touch.

2) I think this defeats the goal of transparency. Users of fully virtualized VMs for instances, have no annotations like that (of course it is possible to hint they are virtualized from many other sources). But in the end of the day, a resource is a resource, virtualized or not.

>
>> 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
>>
> I think some kind of care for users are required as I wrote above.
>

Please note again that I don't necessarily dislike the idea of a mount option, if we must do something. I just don't see the point.

>> 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:
>>
>
> If virtualized /proc/stat works, I don't think 'account only' cgroup is
> necessary. It can be obsolete.
>

Let's keep in mind that there are more to the story, and I want to be sure to address everyones PoVs here. The impression I've got is that the reasons to keep cpuacct around came mainly from 2 sources:

1) Balbir wants statistics that don't interfere with scheduling decisions
2) Paul thinks we should avoid cluttering cpu cgroup.

Obsoleting cpuacct clearly makes somethings simpler, but can probably defeat some of those goals. So still need to hear about this...