Subject: Re: [PATCH v3 04/13] kmem accounting basic infrastructure
Posted by Michal Hocko on Thu, 27 Sep 2012 12:54:15 GMT
View Forum Message <> Reply to Message

On Thu 27-09-12 16:40:03, Glauber Costa wrote:
> On 09/27/2012 04:40 PM, Michal Hocko wrote:
> > On Thu 27-09-12 16:20:55, Glauber Costa wrote:
> >> On 09/27/2012 04:15 PM, Michal Hocko wrote:
> >>> On Wed 26-09-12 16:33:34, Tejun Heo wrote:
> >>> [...]
> >>>>>> So, this seems properly crazy to me at the similar level of
> >>>>>> use_hierarchy fiasco. I'm gonna NACK on this.
> >>>>>
> >>>>> As I said: all use cases I particularly care about are covered by a
> >>>>> global switch.
> >>>>>
> >>>>> I am laying down my views because I really believe they make more sense.
> >>>>> But at some point, of course, I'll shut up if I believe I am a lone voice.
> >>>>>
> >>>>> I believe it should still be good to hear from mhocko and kame, but from
> >>>>> your point of view, would all the rest, plus the introduction of a
> >>>>> global switch make it acceptable to you?
> >>>>
> >>>> The only thing I'm whining about is per-node switch + silently
> >>>> ignoring past accounting, so if those two are solved, I think I'm
> >>>> pretty happy with the rest.
> >>>
> >>> I think that per-group "switch" is not nice as well but if we make it
> >>> hierarchy specific (which I am proposing for quite some time) and do not
> >>> let enable accounting for a group with tasks then we get both
> >>> flexibility and reasonable semantic. A global switch sounds too coars to
> >>> me and it really not necessary.
> >>>
> >>> Would this work with you?
> >>>
> >>
> >> How exactly would that work? AFAIK, we have a single memcg root, we
> >> can't have multiple memcg hierarchies in a system. Am I missing something?
> >
> > Well root is so different that we could consider the first level as the
> > real roots for hierarchies.
> >
> So let's favor clarity: What you are proposing is that the first level
> can have a switch for that, and the first level only. Is that right ?

I do not want any more switches. I am fine with your "set the limit and
start accounting apprach" and then inherit the _internal_ flag down the
hierarchy.

If you are in a child and want to set the limit then you can do that only if your parent is accounted already (so that you can have your own limit). We will need the same thing for oom_controll and swappinness.

--
Michal Hocko
SUSE Labs