

Hello, Glauber.

On Thu, Sep 27, 2012 at 10:45:01PM +0400, Glauber Costa wrote:

> > Can you please give other examples of cases where this type of issue
> > exists (plenty of shared kernel data structure which is inherent to
> > the workload at hand)? Until now, this has been the only example for
> > this type of issues.

>

> Yes. the namespace related caches (*), all kinds of sockets and network
> structures, other file system structures like file struct, vm areas, and
> pretty much everything a full container does.

>

> (*) we run full userspace, so we have namespaces + cgroups combination.

This is probably me being dumb but wouldn't resources used by full namespaces be mostly independent? Which parts get shared? Also, if you do full namespace, isn't it more likely that you would want fuller resource isolation too?

> >> Mel suggestion of not allowing this to happen once the cgroup has tasks
> >> takes care of this, and is something I thought of myself.

> >

> > You mean Michal's? It should also disallow switching if there are
> > children cgroups, right?

>

> No, I meant Mel, quoting this:

>

> "Further I would expect that an administrator would be aware of these
> limitations and set kmem_accounting at cgroup creation time before any
> processes start. Maybe that should be enforced but it's not a
> fundamental problem."

>

> But I guess it is pretty much the same thing Michal proposes, in essence.

>

> Or IOW, if your concern is with the fact that charges may have happened
> in the past before this is enabled, we can make sure this cannot happen
> by disallowing the limit to be set if currently unset (value changes are
> obviously fine) if you have children or any tasks already in the group.

Yeah, please do that.

Thanks.

--

