Subject: Re: [PATCH v3 13/28] slub: create duplicate cache
Posted by Glauber Costa on Tue, 29 May 2012 19:40:02 GMT
View Forum Message <> Reply to Message

On 05/29/2012 11:26 PM, Christoph Lameter wrote:
> On Tue, 29 May 2012, Glauber Costa wrote:
>
>> But we really need a page to be filled with objects from the same cgroup, and
>> the non-shared objects to be accounted to the right place.
>
> No other subsystem has such a requirement. Even the NUMA nodes are mostly
> suggestions and can be ignored by the allocators to use memory from other
> pages.

Of course it does. Memcg itself has such a requirement. The collective
set of processes needs to have the pages it uses accounted to it, and
never go over limit.

>> Otherwise, I don't think we can meet even the lighter of isolation guarantees.
>
> The approach works just fine with NUMA and cpusets. Isolation is mostly
> done on the per node boundaries and you already have per node statistics.

I don't know about cpusets in details, but at least with NUMA, this is
not an apple-to-apple comparison. a NUMA node is not meant to contain
you. A container is, and that is why it is called a container.

NUMA just means what is the *best* node to put my memory.
Now, if you actually say, through you syscalls "this is the node it
should live in", then you have a constraint, that to the best of my
knowledge is respected.

Now isolation here, is done in the container boundary. (cgroups, to be
generic).