
Subject: Re: [PATCH v3 04/13] kmem accounting basic infrastructure
Posted by [Mel Gorman](#) on Thu, 27 Sep 2012 14:43:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Thu, Sep 27, 2012 at 07:33:00AM -0700, Tejun Heo wrote:

> Hello, Michal.

>

> On Thu, Sep 27, 2012 at 02:08:06PM +0200, Michal Hocko wrote:

> > Yes, because we have many users (basically almost all) who care only

> > about the user memory because that's what occupies the vast majority of

> > the memory. They usually want to isolate workload which would disrupt

> > the global memory otherwise (e.g. backup process vs. database). You

> > really do not want to pay an additional overhead for kmem accounting

> > here.

>

> I'm not too convinced. First of all, the overhead added by kmemcg

> isn't big.

Really?

If kmemcg was globally accounted then every `__GFP_KMEMCG` allocation in the page allocator potentially ends up down in `__memcg_kmem_newpage_charge` which

1. takes RCU read lock
2. looks up cgroup from task
3. takes a reference count
4. `memcg_charge_kmem` -> `__mem_cgroup_try_charge`
5. release reference count

That's a *LOT* of work to incur for cgroups that do not care about kernel accounting. This is why I thought it was reasonable that the kmem accounting not be global.

--

Mel Gorman
SUSE Labs
