On Tue, 16 Oct 2012, Glauber Costa wrote:

> From: Suleiman Souhlal <ssouhlal@FreeBSD.org>
>
> mem_cgroup_do_charge() was written before kmem accounting, and expects
> three cases: being called for 1 page, being called for a stock of 32
> pages, or being called for a hugepage.  If we call for 2 or 3 pages (and
> both the stack and several slabs used in process creation are such, at
> least with the debug options I had), it assumed it's being called for
> stock and just retried without reclaiming.
>
> Fix that by passing down a minsize argument in addition to the csize.
>
> And what to do about that (csize == PAGE_SIZE && ret) retry?  If it's

I think you're referring to the (nr_pages == 1 && ret) retry, csize is
only used for interfacing with res_counter.

> needed at all (and presumably is since it's there, perhaps to handle
> races), then it should be extended to more than PAGE_SIZE, yet how far?
> And should there be a retry count limit, of what?  For now retry up to
> COSTLY_ORDER (as page_alloc.c does) and make sure not to do it if
> __GFP_NORETRY.
>
> [v4: fixed nr pages calculation pointed out by Christoph Lameter ]
>
> Signed-off-by: Suleiman Souhlal <suleiman@google.com>
> Signed-off-by: Glauber Costa <glommer@parallels.com>
> Acked-by: Kamezawa Hiroyuki <kamezawa.hiroyu@jp.fujitsu.com>
> Acked-by: Michal Hocko <mhocko@suse.cz>
> Acked-by: Johannes Weiner <hannes@cmpxchg.org>
> CC: Tejun Heo <tj@kernel.org>

Acked-by: David Rientjes <rientjes@google.com>