
Subject: Re: [PATCH 6/6 mm] memcgroup: revert swap_state mods
Posted by [Balbir Singh](#) on Mon, 12 Nov 2007 06:56:04 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hugh Dickins wrote:

> If we're charging rss and we're charging cache, it seems obvious that
> we should be charging swapcache - as has been done. But in practice
> that doesn't work out so well: both swapin readahead and swapoff leave
> the majority of pages charged to the wrong cgroup (the cgroup that
> happened to read them in, rather than the cgroup to which they belong).
>
> (Which is why unuse_pte's GFP_KERNEL while holding pte lock never
> showed up as a problem: no allocation was ever done there, every page
> read being already charged to the cgroup which initiated the swapoff.)
>
> It all works rather better if we leave the charging to do_swap_page and
> unuse_pte, and do nothing for swapcache itself: revert mm/swap_state.c
> to what it was before the memory-controller patches. This also speeds
> up significantly a contained process working at its limit: because it
> no longer needs to keep waiting for swap writeback to complete.
>

Yes, it does speed up things, but we lose control over swap cache.
It might grow very large, but having said that I am in favour of
removing the mods till someone faces a severe problem with them.
Another approach is to provide a per-container tunable as to
whether swap cache should be controlled or not and document
the side-effects of swap cache control.

> Is it unfair that swap pages become uncharged once they're unmapped,
> even though they're still clearly private to particular cgroups? For
> a short while, yes; but PageReclaim arranges for those pages to go to
> the end of the inactive list and be reclaimed soon if necessary.
>
> shmem/tmpfs pages are a distinct case: their charging also benefits
> from this change, but their second life on the lists as swapcache
> pages may prove more unfair - that I need to check next.
>
> Signed-off-by: Hugh Dickins <hugh@veritas.com>

Thanks for the patch

Acked-by: Balbir Singh <balbir@linux.vnet.ibm.com>

--
Warm Regards,
Balbir Singh

Linux Technology Center
IBM, ISTL

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>
