Subject: Re: [PATCH 0/5] blk-throttle: writeback and swap IO control
Posted by Balbir Singh on Thu, 24 Feb 2011 06:08:53 GMT
View Forum Message <> Reply to Message

* Andrea Righi <arighi@develer.com> [2011-02-22 18:12:51]:

> Currently the blkio.throttle controller only support synchronous IO requests.
> This means that we always look at the current task to identify the "owner" of
> each IO request.
>
> However dirty pages in the page cache can be wrote to disk asynchronously by
> the per-bdi flusher kernel threads or by any other thread in the system,
> according to the writeback policy.
>
> For this reason the real writes to the underlying block devices may
> occur in a different IO context respect to the task that originally
> generated the dirty pages involved in the IO operation. This makes the
> tracking and throttling of writeback IO more complicate respect to the
> synchronous IO from the blkio controller's perspective.
>
> The same concept is also valid for anonymous pages involed in IO operations
> (swap).
>
> This patch allow to track the cgroup that originally dirtied each page in page
> cache and each anonymous page and pass these informations to the blk-throttle
> controller. These informations can be used to provide a better service level
> differentiation of buffered writes swap IO between different cgroups.
>
> Testcase
> ========
> - create a cgroup with 1MiB/s write limit:
> # mount -t cgroup -o blkio none /mnt/cgroup
> # mkdir /mnt/cgroup/foo
> # echo 8:0 $((1024 * 1024)) > /mnt/cgroup/foo/blkio.throttle.write_bps_device
>
> - move a task into the cgroup and run a dd to generate some writeback IO
>
> Results:
> - 2.6.38-rc6 vanilla:
> $ cat /proc/$$/cgroup
> 1:blkio:/foo
> $ dd if=/dev/zero of=zero bs=1M count=1024 &
> $ dstat -df
> --dsk/sda--
> read writ
> 0 19M
> 0 19M
> 0 0

> 0 0
> 0 19M
> ...
>
> - 2.6.38-rc6 + blk-throttle writeback IO control:
> $ cat /proc/$$/cgroup
> 1:blkio:/foo
> $ dd if=/dev/zero of=zero bs=1M count=1024 &
> $ dstat -df
> --dsk/sda--
> read writ
> 0 1024
> 0 1024
> 0 1024
> 0 1024
> 0 1024
> ...
>

Thanks for looking into this, further review follows.

--
Three Cheers,
Balbir

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containe rs