
Subject: Re: [PATCH] memory cgroup enhancements take 4 [5/8] add status accounting function for memory cgroup

Posted by [akpm](#) on Wed, 31 Oct 2007 22:12:34 GMT

[View Forum Message](#) <> [Reply to Message](#)

On Wed, 31 Oct 2007 19:30:46 +0900

KAMEZAWA Hiroyuki <kamezawa.hiroyu@jp.fujitsu.com> wrote:

```
> Add statistics account infrastructure for memory controller.
> All account information is stored per-cpu and caller will not have
> to take lock or use atomic ops.
> This will be used by memory.stat file later.
>
> CACHE includes swapcache now. I'd like to divide it to
> PAGECACHE and SWAPCACHE later.
>
> ...
>
> --- devel-2.6.23-mm1.orig/mm/memcontrol.c
> +++ devel-2.6.23-mm1/mm/memcontrol.c
> @@ -35,6 +35,59 @@ struct cgroup_subsys mem_cgroup_subsys;
> static const int MEM_CGROUP_RECLAIM_RETRIES = 5;
>
> /*
> + * Statistics for memory cgroup.
> + */
> +enum mem_cgroup_stat_index {
> + /*
> + * For MEM_CONTAINER_TYPE_ALL, usage = pagecache + rss.
> + */
> + MEM_CGROUP_STAT_CACHE, /* # of pages charged as cache */
> + MEM_CGROUP_STAT_RSS, /* # of pages charged as rss */
> +
> + MEM_CGROUP_STAT_NSTATS,
> +};
> +
> +struct mem_cgroup_stat_cpu {
> + s64 count[MEM_CGROUP_STAT_NSTATS];
> +} ____cacheline_aligned_in_smp;
> +
> +struct mem_cgroup_stat {
> + struct mem_cgroup_stat_cpu cpustat[NR_CPUS];
> +};
> +
> +/*
> + * modifies value with disabling preempt.
> + */
> +static inline void __mem_cgroup_stat_add(struct mem_cgroup_stat *stat,
```

```
> + enum mem_cgroup_stat_index idx, int val)
> +{
> + int cpu = smp_processor_id();
> + preempt_disable();
> + stat->cpustat[cpu].count[idx] += val;
> + preempt_enable();
> +}
```

This is clearly doing `smp_processor_id()` in preemptible code. (or the `preempt_disable()` just isn't needed). I fixed it up.

Please ensure that you test with all runtime debugging options enabled - you should have seen a warning here.

```
> +/*
> + * For accounting under irq disable, no need for increment preempt count.
> + */
> +static inline void __mem_cgroup_stat_add_safe(struct mem_cgroup_stat *stat,
> + enum mem_cgroup_stat_index idx, int val)
> +{
> + int cpu = smp_processor_id();
> + stat->cpustat[cpu].count[idx] += val;
> +}
```

There's a wild amount of inlining in that file. Please, just don't do it - inline is a highly specialised thing and is rarely needed.

When I removed the obviously-wrong inline statements, the size of `mm/memcontrol.o` went from 3823 bytes down to 3495.

It also caused this:

```
mm/memcontrol.c:65: warning: '__mem_cgroup_stat_add' defined but not used
```

so I guess I'll just remove that.

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