Subject: [PATCH v3 2/6] account guest time per-cgroup as well.
Posted by Glauber Costa on Wed, 30 May 2012 09:48:33 GMT
View Forum Message <> Reply to Message

We already track multiple tick statistics per-cgroup, using
the task_group_account_field facility. This patch accounts
guest_time in that manner as well.

Signed-off-by: Glauber Costa <glommer@parallels.com>
CC: Peter Zijlstra <a.p.zijlstra@chello.nl>
CC: Paul Turner <pjt@google.com>
---
 kernel/sched/core.c | 10 ++++------
 1 file changed, 4 insertions(+), 6 deletions(-)

diff --git a/kernel/sched/core.c b/kernel/sched/core.c
index 39eb601..220d416 100644
--- a/kernel/sched/core.c
+++ b/kernel/sched/core.c
@@ -2717,8 +2717,6 @@ void account_user_time(struct task_struct *p, cputime_t cputime,
 static void account_guest_time(struct task_struct *p, cputime_t cputime,
 cputime_t cputime_scaled)
 {
- u64 *cpustat = kcpustat_this_cpu->cpustat;
-
 /* Add guest time to process. */
 p->utime += cputime;
 p->utimescaled += cputime_scaled;
@@ -2727,11 +2725,11 @@ static void account_guest_time(struct task_struct *p, cputime_t
cputime,

 /* Add guest time to cpustat. */
 if (TASK_NICE(p) > 0) {
- cpustat[CPUTIME_NICE] += (__force u64) cputime;
- cpustat[CPUTIME_GUEST_NICE] += (__force u64) cputime;
+ task_group_account_field(p, CPUTIME_NICE, (__force u64) cputime);
+ task_group_account_field(p, CPUTIME_GUEST, (__force u64) cputime);
 } else {
- cpustat[CPUTIME_USER] += (__force u64) cputime;
- cpustat[CPUTIME_GUEST] += (__force u64) cputime;
+ task_group_account_field(p, CPUTIME_USER, (__force u64) cputime);
+ task_group_account_field(p, CPUTIME_GUEST, (__force u64) cputime);
 }
 }

--
1.7.10.2

Subject: Re: [PATCH v3 2/6] account guest time per-cgroup as well.
Posted by Peter Zijlstra on Wed, 30 May 2012 10:32:27 GMT
View Forum Message <> Reply to Message

On Wed, 2012-05-30 at 13:48 +0400, Glauber Costa wrote:
> We already track multiple tick statistics per-cgroup, using
> the task_group_account_field facility. This patch accounts
> guest_time in that manner as well.

So this leaves IOWAIT and IDLE the only fields not using
task_group_account_field(), right?

Subject: Re: [PATCH v3 2/6] account guest time per-cgroup as well.
Posted by Glauber Costa on Wed, 30 May 2012 10:36:27 GMT
View Forum Message <> Reply to Message

On 05/30/2012 02:32 PM, Peter Zijlstra wrote:
> On Wed, 2012-05-30 at 13:48 +0400, Glauber Costa wrote:
>> > We already track multiple tick statistics per-cgroup, using
>> > the task_group_account_field facility. This patch accounts
>> > guest_time in that manner as well.
> So this leaves IOWAIT and IDLE the only fields not using
> task_group_account_field(), right?

Yes, because they are essentially global, and their meaning is
ill-defined from within a cgroup.

If you look further out in the patchset, I intend to export idle from
cpu, instead of cpuacct, because something that can be used as idle
value is already computed anyway from the schedstats, so I'm just using it.

iowait will be left blank for now. Me and Paul agreed last time we
talked that it is not uber important to have iowait values per-cgroup.

Subject: Re: [PATCH v3 2/6] account guest time per-cgroup as well.
Posted by Paul Turner on Wed, 30 May 2012 10:46:13 GMT
View Forum Message <> Reply to Message

On Wed, May 30, 2012 at 3:36 AM, Glauber Costa <glommer@parallels.com> wrote:
> On 05/30/2012 02:32 PM, Peter Zijlstra wrote:
>>
>> On Wed, 2012-05-30 at 13:48 +0400, Glauber Costa wrote:
>>>
>>> > We already track multiple tick statistics per-cgroup, using
>>> > the task_group_account_field facility. This patch accounts
>>> > guest_time in that manner as well.

>>
>> So this leaves IOWAIT and IDLE the only fields not using
>> task_group_account_field(), right?
>
>
> Yes, because they are essentially global, and their meaning is ill-defined
> from within a cgroup.
>
> If you look further out in the patchset, I intend to export idle from cpu,
> instead of cpuacct, because something that can be used as idle value is
> already computed anyway from the schedstats, so I'm just using it.
>
> iowait will be left blank for now. Me and Paul agreed last time we talked
> that it is not uber important to have iowait values per-cgroup.

Stronger: it lacks a definition you can sanely measure without atomic
counters everywhere (similarly for group-idle).

> --
> To unsubscribe from this list: send the line "unsubscribe cgroups" in
> the body of a message to majordomo@vger.kernel.org
> More majordomo info at http://vger.kernel.org/majordomo-info.html