Subject: Re: [RFC] [PATCH 0/3] Add group fairness to CFS
Posted by Peter Williams on Thu, 24 May 2007 03:15:04 GMT
View Forum Message <> Reply to Message

Srivatsa Vaddagiri wrote:
> Here's an attempt to extend CFS (v13) to be fair at a group level, rather than
> just at task level. The patch is in a very premature state (passes
> simple tests, smp load balance not supported yet) at this point. I am sending
> it out early to know if this is a good direction to proceed.
>
> Salient points which needs discussion:
>
> 1. This patch reuses CFS core to achieve fairness at group level also.
>
>   To make this possible, CFS core has been abstracted to deal with generic
>   schedulable "entities" (tasks, users etc).
>
> 2. The per-cpu rb-tree has been split to be per-group per-cpu.
>
>   schedule() now becomes two step on every cpu : pick a group first (from
>   group rb-tree) and a task within that group next (from that group's task
>   rb-tree)
>
> 3. Grouping mechanism - I have used 'uid' as the basis of grouping for
>   timebeing (since that grouping concept is already in mainline today).
>   The patch can be adapted to a more generic process grouping mechanism
>   (like http://lkml.org/lkml/2007/4/27/146) later.
>
> Some results below, obtained on a 4way (with HT) Intel Xeon box. All
> number are reflective of single CPU performance (tests were forced to
> run on single cpu since load balance is not yet supported).
>
>
>   uid "vatsa"   uid "guest"
>   (make -s -j4 bzImage)   (make -s -j20 bzImage)
>
> 2.6.22-rc1   772.02 sec 497.42 sec (real)
> 2.6.22-rc1+cfs-v13   780.62 sec 478.35 sec (real)
> 2.6.22-rc1+cfs-v13+this patch   776.36 sec 776.68 sec (real)

This would seem to indicate that being fair between groups isn't always
a good thing. With 2.6.22-rc1 and 2.6.22-rc1+cfs-v13 "guest" gets his
build done in about 2/3 the time of "vatsa" without seriously
inconveniencing "vatsa". All making scheduling fair between the groups
has done is penalize "guest" without significantly improving matters for
"vatsa" (he gains a mere 4 seconds out of 780).

BUT I imagine that this is an artefact caused by the use of HT

technology and that if the test were run on a computer without HT the results would be more impressive.

Peter
--
Peter Williams   pwil3058@bigpond.net.au

"Learning, n. The kind of ignorance distinguishing the studious."
  -- Ambrose Bierce

_____

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