Rework load_balance

Vincent Guittot
Linux Plumber Conference
Scheduler MC
9 September 2019
Status

- Patchset on LKML
  - v3 still under preparation

- Changes:
  - Extend groups classification
  - Define different type of migration
  - Better take into account new metrics & remove old heuristics
  - Fix some suboptimal tasks placements
  - Use load_avg instead of runnable_load_avg

- Among fixed UCs:
  - 1 task per CPU
  - Better spread tasks in cgroup on numa
  - Preemption by other classes
Open item #1 runnable load vs load

- Runnable load introduced to
  - Fix cases with huge blocked load

- Load balance
  - Replaced runnable_load_avg by load_avg

- Wakeup path uses runnable load
  - Align policy with load_balance
  - A proposal submitted to use on load_avg but need refinements

- Numa stats uses runnable_load too
  - Study the impact of aligning with load_balance()
Open item #2: detection of overloaded state

- How to better detect overloaded CPU/groups?
- $\text{util}_{\text{avg}}/\text{util}_{\text{est}}$ can be temporarily low after migration
Open item #2: detection of overloaded state

- But load and runnable load stay quite high
  - tasks @ root with nice 0

- Nice weight screws up this information
### Open item #2: detection of overloaded state

- Example of load balance stats for hackbench running on a dual quad core

<table>
<thead>
<tr>
<th></th>
<th>mainline</th>
<th>patchset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nr=0</td>
<td>nr=1</td>
</tr>
<tr>
<td>Total</td>
<td>1800</td>
<td>312</td>
</tr>
<tr>
<td>has_capacity</td>
<td>633</td>
<td>35%</td>
</tr>
<tr>
<td>fully_busy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>misfit</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>asym Packing</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Overloaded</td>
<td>1167</td>
<td>65%</td>
</tr>
</tbody>
</table>

Note: The patchset has_capacity column shows a significant increase compared to the mainline, indicating potential issues with resource allocation.
Open item #3 fairness

- How to ensure better fairness
  - N+1 tasks on N CPUs case
- What drives the migration?
  - `nr_balance_failed`
  - `load / 2 <= env->imbalance`
Thank you

Join Linaro to accelerate deployment of your Arm-based solutions through collaboration

contact@linaro.org