Recent Power Management Enhancements in DPDK

DAVE HUNT, CHRIS MACNAMARA, LIANG MA, RADU NICOLAU
Updates Since Last Time

• Quick reminder…. feature review

• Updates & Discussions to follow
  
  • How Busy Am I? 100% …. Hmmm

  • New Methods to Trigger Power State Changes
    • Load conditions in a 100% Polling environment
    • Out of Band Energy Efficiency determination for 100% Polling DPDK PMDs

  • Power Policies for Containers
Existing DPDK Power Capabilities

Many use cases, support for direct control, virtualized architecture
## Existing DPDK Power Features

<table>
<thead>
<tr>
<th>Challenge / Problem</th>
<th>DPDK Solution / Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>L3fwd power using C states</td>
<td>Sample app</td>
</tr>
<tr>
<td>Traffic always running, always on cores</td>
<td>Added core Frequency State APIs</td>
</tr>
<tr>
<td>Increase performance on key cores when busy or overloaded</td>
<td>Added Turbo Boost APIs in rte_power.h</td>
</tr>
<tr>
<td>Virtualized Software Architecture: Long latency of a VM detect, waste of monitoring and changing state, move to policy based control</td>
<td>Inband: New SW Arch for policy control via virtio-serial</td>
</tr>
<tr>
<td>Match CPU power to network load (Scale down when not busy, turbo when busy)</td>
<td>Sample app: Time of day</td>
</tr>
<tr>
<td>Fast scale up when burst arrives</td>
<td>Sample app: Packet arrival rate (NIC stats)</td>
</tr>
</tbody>
</table>

Librte_power APIs and Sample Apps
New DPDK Features Since Last Time!

<table>
<thead>
<tr>
<th>Challenge / Problem</th>
<th>DPDK Solution / Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pin DPDK threads/lcores to high priority cores</td>
<td>Pinning relevant workloads to Turbo Cores</td>
</tr>
<tr>
<td>App Agnostic mechanism to detect when DPDK is 100% polling and no packets or work</td>
<td>Sample code: Branch prediction ratio used as trigger to detect idle -&gt; modify power</td>
</tr>
<tr>
<td>Mechanism to determine load (Experimental branch)</td>
<td>Empty polling trend analysis and trigger to modify power (e.g. how busy)</td>
</tr>
<tr>
<td>Power Policies for Containers</td>
<td>New FIFO interface to Power Manager that accepts policies via JSON</td>
</tr>
</tbody>
</table>

New triggers and capabilities enabling new use cases
Determining Load
How busy am I ... determining load

Using empty polls useful for load detection & action trigger
Pushed Patches To Support This (Traffic Aware)

- Submitted to mailing list
  http://patches.dpdk.org/project/dpdk/list/?series=1143
- API marked as experimental

```
@Init
How many polls can we do
Set thresholds (idle/busy)
```

```
@run
Count empty polls
Check against threshold
Am I busy?
```

```
@run
Adjust State, snooze or run faster
```
Out of Band
Energy Efficiency
Poll Loop Work Rate Detection (PMD Load%)

- CPU Load is always 100% for DPDK PMD Poll Loops
- Actual workload may be zero (processing zero packets)
- Use the ratio between Branch Hits and Branch Misses
- Ratio is low when tight code loop (empty polling), and significantly is higher when processing packets (due to larger code path)
- *Almost* linear with traffic rate

Application Agnostic Idle Detection using Branching
Latency Comparison

- Using out-of-band branch ratio mechanism of power management (Merged in 18.08)

- Three measurements shown
  - 2.5GHz fixed core frequency
  - 1.0GHz fixed core frequency
  - 1.0GHz – 2.5GHz variable base on branch ratop

- Branch Ratio mechanism reading core counters every 100uS

Branch Ratio Mechanism Latency Comparison

Branch Ratio Latency as expected
Pushed Patches To Support This (Branch Ratio)

- Applied in DPDK 18.08
Power Policies for Containers
Current mechanism to send policies from VMs to Power Manager via virtop-serial

New patch-set adds additional interface into Power Manager via file system FIFO

Handles existing power commands, max, min, up, down, etc.

Handles power polices, similar to VM virtio-serial channels

Can be used by any application with access to the FIFO in the Host OS
- Host Applications
- Container Applications
Pushed Patches To Support This (Policies for Containers)

- Submitted to mailing list for 18.11
- http://patches.dpdk.org/project/dpdk/list/?series=1109
Thank You

Chris MacNamara (chris.macnamara@intel.com)
Dave Hunt (david.hunt@intel.com)
Liang Ma <liang.j.ma@intel.com>

Acknowledgements
Mike Glynn, John Geary, Stephen Byrne, Tim O’Driscoll, Walt Gilmore