OPEN CLOUD SERVER
PROJECT OLYMPUS
Power Capping

Ali Larijani
Microsoft
F/W Engineering Manager

David Locklear
Intel
Platform Architect
Power Capping

Agenda:

• Power capping benefits at Data center
• Power Capping methods (Static vs. Dynamic)
• Intel Power Node Manager
• Project Olympus Power capping
• Power capping Examples
What is Power capping?

- Power Capping is a technique to keep power consumption below a threshold without any interruption to server operation.
- Power Capping can be hierarchically applied at server, rack, rows,...
Total cost of power & cooling comes as second greatest operating cost after Servers cost so power efficiency should be highly invested

**Benefits of Power Capping:**
- Operation continuity is improved by limiting H/W overheat and fail
- Performance/Watt and efficiency can be increased
- Stranded power can be reduced
- Allows dynamic balancing of power and cooling resources by moving them to demanding workloads
Power Capping

Power Capping method (Static vs. Dynamic)

- **Static**
  - Fixed per name plate data but with low power utilization
  - Always active – policy in effect

- **Dynamic**
  - Power steered to servers w/greater workload
  - Can be Adaptive with workload variations
  - Policy applied whenever needed
Intel® Node Manager - Power Control

- Monitors server power with short control loop to HSC
- Current power level continuously compared to policy
- FAST Limiting - Closed loop process adjusts CPU & Memory power
- No impact under normal operation – Server performs as much work as possible with restricted power
- Node Throttling for fast response to power delivery issues - Short duration prevents impact to workloads.
Project Olympus Power capping
Project Olympus Power capping

Power Capping Triggers

1) Server Level:
   - **Fast Proc Hot**: asserted for a minimum 100ms under OC(> 65A) and UV(<11.5V)
   - **VR Hots**: CPU VRs can also generate Proc Hot triggers to CPUs
   - **PSU Alert**: PSU has N+2 design including x3 340W redundant modules
     - Triggered when x2 modules failed with OC condition exists on third module
     - PSU alert assertion limits server’s power to 340W

*PSU = Input Voltage Selector*
Project Olympus Power capping

Power Capping Triggers

2) Rack Level:

- Power monitoring is continuously running at rack level
- If rack power consumption exceeds threshold, RM Throttle is asserted
- RM can set a policy of “Not Action”, “DPC only” or “DPC + Proc Hot” per server
Power Capping

Power capping examples:

Improving Server Availability:
• PSU power capacity of 1040w reduced to 340W due to power module failures, Server power consumption capped to 340W to ensure operation continuity

Improving Rack Density
• 14000W compute power = 28 nodes at 500W (TDP) or 32 nodes at 435W (capped)

Improving Power Utilization
• steering power to servers with higher work loads:
  • 14000W compute power = 12 nodes at full 500W, 20 nodes at 400W
RM/BMC /Intel ME interactions to Realize a dynamic power cap

- Upon RM_Throttle assertion:
  - CPUs throttled by going into LFM mode
  - BMC send DPC level to Intel ME
  - BMC force Proc Hot release
  - Intel ME runs a power control loop
  - HSC power ram-up monotonically to DPC level