CERN openlab

Intel Data Center Manager

Julien Leduc
CERN openlab Intel Fellow
julien.leduc at cern.ch

Introduction

- Intel Data Center Manager is a web service aimed at managing Data Center management, built on top of:
  - Apache tomcat
  - PostgreSQL database
- Monitors two metrics on the managed entities:
  - Node power consumption
  - Inlet temperature
- Groups managed entities in hierarchies

- One can then defines policies to Groups or individual servers to cap the power

Managed Entities

- Servers equipped with one of the following technologies:
  - Intel Power Node Manager (v1.5 or v2.0)
  - Dell iDRAC 6 Server
    - Integrated Dell Remote Access Controller 6
  - Data Center Manageability Interface

- IPNM and iDRAC6 are IPMI 2.0 compliant, but using specific OEM functions

Groups

- Allows grouping of monitored entities
  - Datacenter, room, row, rack, logical group
- Specify location of this group
- Defines a power limit to the group, for example:
  - Breakers for a rack
  - PDU for a row
- Allows adding “unmanaged equipment power”: for example:
  - Switch for a rack
  - KVM, screen for a row

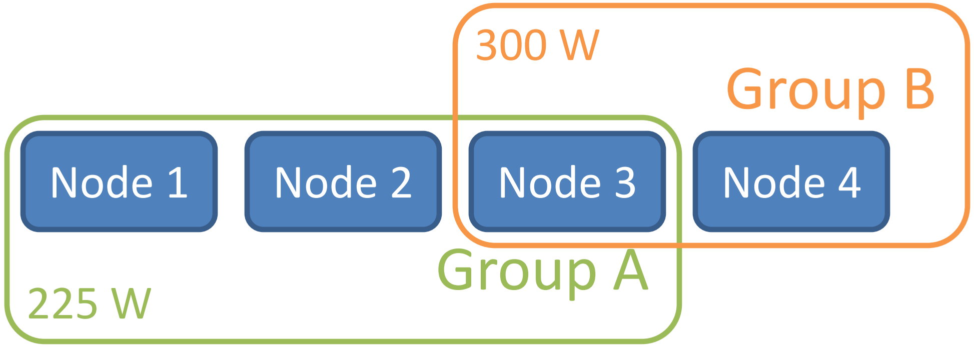
Defining power policies

- One can define some policies and apply them to managed entities or groups
- Several policies can be defined for a managed entity thanks to priorities
  - High
  - Medium (default)
  - Low
- DCM then tries to maximize the power allocated to High priority

Power limiting policies

- There are different types of policies:
  - Custom power limit: limits the power consumption of an entity
  - Minimum power: provides the minimum possible power to an entity
  - Minimum power on inlet temperature trigger: provides the minimum possible power to an entity if inlet temperature is higher than a defined value
- These policies can be scheduled, and prioritized

Policies example



Where Group A priority is higher that Group B

Each node in Group A receives 75W.

For Group B, the total power supplied is 300w. Since Group A limits Node3 to 75W, Node4 receives 225W.

If the power policy for Group A changes, then the implementation of the power policy for Group B could change also, even if the Group B policy itself does not change.

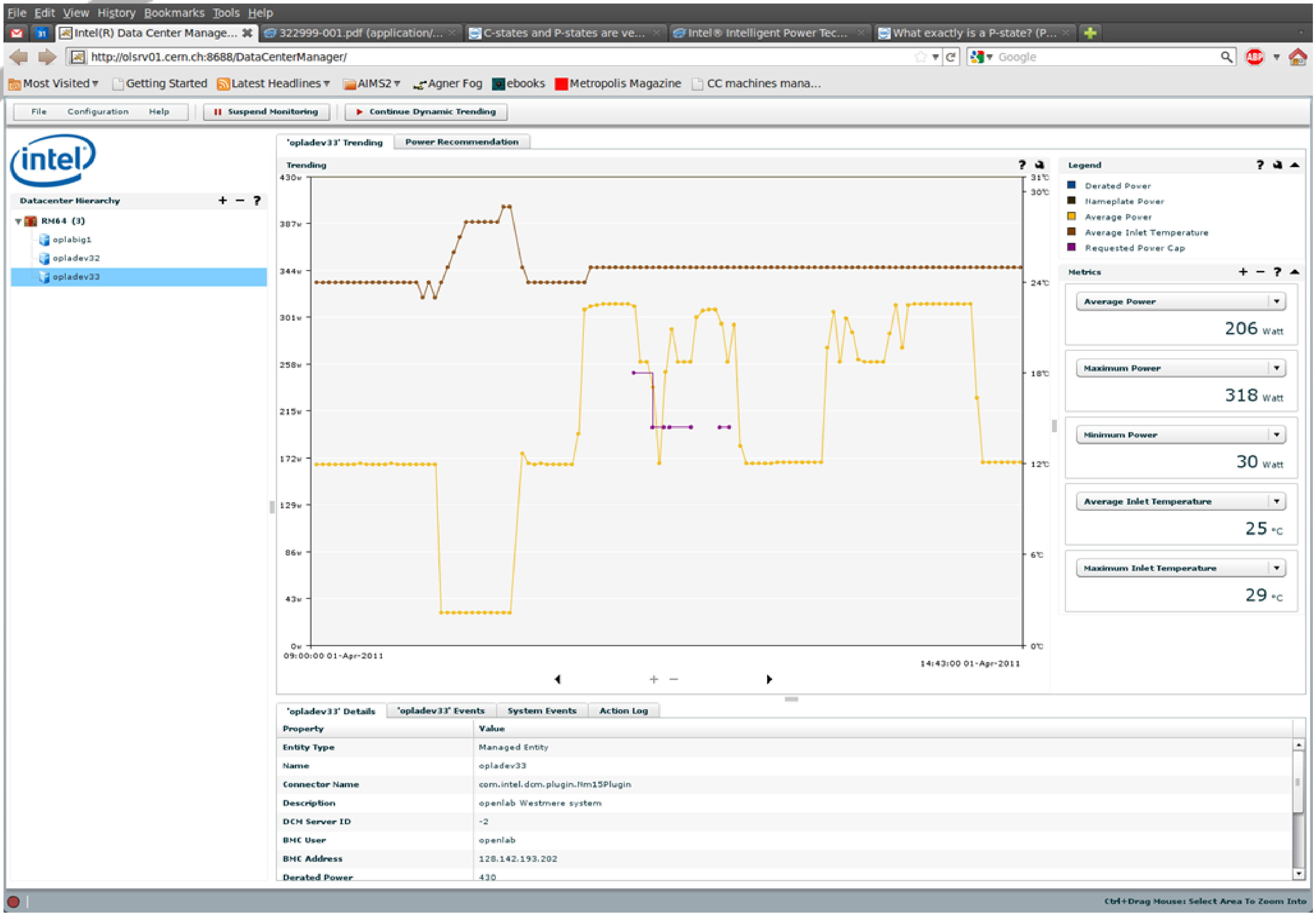
Applying power policies

- Prerequisites for the managed entities specify that the OS must be ACPI compliant
- Intel IPNM initiates P-states changes through ACPI
  - Allowing then the OS to control the CPUs voltage and frequency

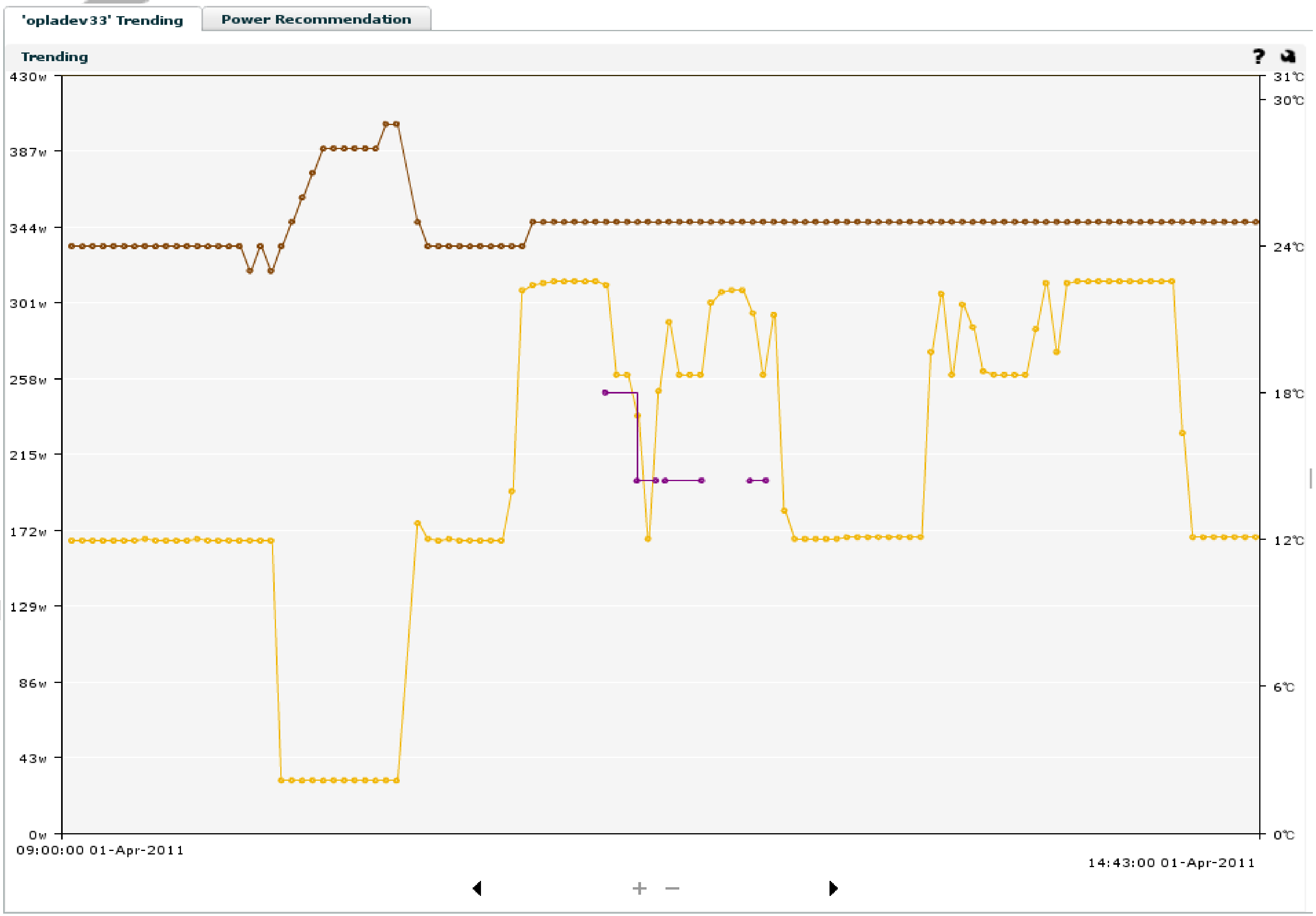
Policy enabled:

Cn	Avg residency	P-states (freq)	Avg residency
C0	100%	Turbo Mode	0%
C1	0%	2.27 GHz	0%
C2	0%	2.13 GHz	0%
C3	0%	2.00 GHz	0%
		1.60 GHz	100%

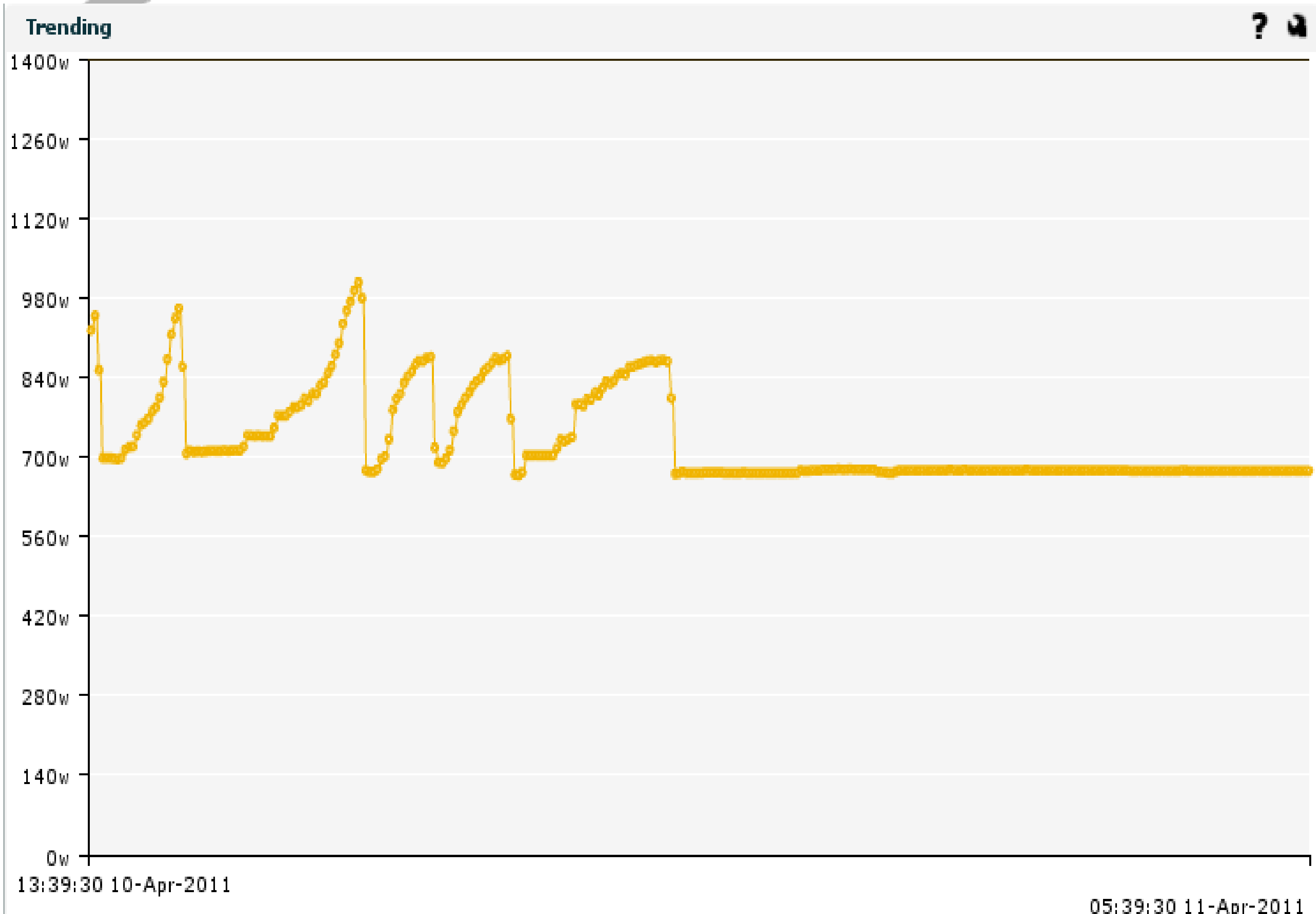
DCM web interface



DCM policy example



DCM monitoring example



Limitations

- Measured power is active power and not apparent power
  - Need to profile the PSU and power factor in for every type of server to deduce apparent power (information could be added in CDB)

Conclusion

- DCM is interesting, using standard technologies
- But IPNM is even more interesting, allowing to gather additional metrics and program policies directly on the node
- Additional work needed:
  - Direct IPMI communication with IPNM, to control the node without DCM
- CERN's production server are not ready
  - Experiment on some Dell servers using lemon to log power consumption