SCE DRET Findings
JUNE 2017

DR13.08: EPRI EB III D – Automated Demand Response in Data Centers

SERVER POWER MANAGEMENT DECREASES ENERGY CONSUMPTION

Data centers must operate 24/7 and exhibit near-constant load on the grid, with relatively flat demand profiles in terms of daily and seasonal changes. In fact, data centers can consume more than 40 times the power density of conventional office space, making it one of the most energy-intense commercial/industrial building types . While data centers have been identified to be a good candidate for energy efficiency improvements, they been remained a relatively untapped market for Demand Response (DR) programs which is due to the critical services data centers support and the potential unplanned downtime costs that can occur due to power outages. Southern California Edison (SCE) in partnership with Electric Power Research Institute (EPRI) and the University of California, Irvine (UCI) have set out to study laboratory and real-world use of server power management systems in response to DR events. The technology selected for this demonstration is the power-capping feature in servers using the latest line of Intel® processors, enabled by Intel's Node Manager technology. This study seeks to better understand the capabilities of this feature in terms of power reduction, time to respond, impact to operations, and post-event rebound, as well as to show the ability for this feature to be initiated as an automated response to an external DR signal.

INTRODUCTION

What Is This Technology?

SERVER MANAGEMENT SOFTWARE FOR DATA CENTER DR OPPORTUNITIES

Power capping in data centers is achieved through Data Center Infrastructure Management (DCIM) tools. The server reports power and temperature over standard communications protocols for monitoring with DCIM tools to manage data center assets and operations. In addition to monitoring capabilities, these technologies enable energy consumption management by placing specific power limits on each server. The server hardware responds to these commands by adjusting its operating state in terms of both performance and throttle, manipulating the voltage and frequency of the CPU and introducing a small-time delay between each computation cycle. In this way, the technology is able to place a cap (or limit) on the power draw of a server or group of servers under management.

For this study, EPRI selected Intel Node Manager technology enabled by Intel's Data Center Manger (DCM), a basic DCIM tool to monitor and manage power within the data center. Finally, a custom version of Energy Interop Server & System (EISS) Client software was delivered by IPKeys to provide end-node response to OpenADR (Open Automated Demand Response) signals from a utility or power system operator.

Figure 1: Diagram of Communications Pathway

[1]*National Resources Defense Council (NRDC), "Data Center Efficiency Assessment: Scaling Up Energy Efficiency Across the Data Center Industry: Evaluating Key Drivers and Barriers," New York, NY, 2014.*

What We Did

LABORATORY AND FIELD TESTING

EPRI mirrored testing in a laboratory environment, and in an operational data center at Calit2 at UCI, using five Dell servers equipped with Intel processors compatible with its Node Manager technology.

The study assessed three separate functions of the technology to demonstrate its DR feasibility: the technology should be demonstrated to have the ability to effectively limit server power; the technology should be shown to reduce the average server power use under typical workload; the technology should be able to initiate a power restriction in response to an external signal, showing the feasibility of utilizing this technology to automatically respond without human intervention to a DR signal from a utility or other power system operator.

Table 1: Laboratory and Field Servers used:

Server	Model	CPU TDW	RAM	Drives	Environment	Application(s)
1	R520 (2U)	E5-2430 (95W)	2x 8GB	3x SATA, 2x SAS	VMWare	Web server, data center manager, energy apps
2	R520 (2U)	E5-2450 (95W)	2x 16GB	2x SATA, 3x SAS	Windows	File server
3	R320 (1U)	E5-2407 (80W)	2x 8GB	2x SATA	Windows	Active directory, domain name ser-vice
4	R320 (1U)	E5-2407 (80W)	2x 8GB	2x SATA	Windows	Active directory, domain name ser-vice
5	R320 (1U)	E5-2407 (80W)	2x 8GB	2x SATA	Windows	Misc. applications

DATA MONTIORING PLAN

Data Point Capture & Equipment Used

Electrical Characteristics

- Voltage [volts]
- Current [amps]
- Power [W]
- Power Factor
- Energy Consumption [kWh]
- Total Harmonic Distortion (THD) [% of amps]

Thermal Characteristics

Temperature of air exhausted from servers [°F]; inlet air temperature was available from the server internal monitoring sensors

Data Monitoring Equipment

The equipment used for metering and data collection at the field site is listed below:

- Power meter: Elkor WattsOn (revenue grade)
- Current transformers (CTs): Continental Controls Accu-CT (15 A, revenue grade)
- Sealed temperature sensors: Omega HSTH-44031
- Communications — Obvius products:
  - **Data acquisition unit:** AcquiSuite A8810
  - **Input/output module:** Flex IO remote

POWER CAPPING TO SIGNIFICANT LEVELS MAY BE CHALLENGING TO REALIZE Field testing results indicate that power capping successfully reduced the average power of the web server under load by as much as 17.9% (18.2W) from the baseline. However, the average time needed to complete HTTP requests (including latency and length of response) under each power cap increased. At the lowest power caps, the time to deliver HTTP requests increased by over 4300%, which suggests that power capping to such significant levels may not be feasible except in extreme situations.

AVERAGE SERVER POWER UNDER POWER CAPPING FOR FIELD WEB SERVER TEST 1 (5 USERS)

Power Cap	Baseline	110 W	100 W	90W	80W	Minimum
Average Power	101.9 W	102.3 W	97.0 W	88.4 W	83.7 W	83.7 W
Savings %	N/A	-0.40%	4.80%	13.20%	17.90%	17.90%
HTTP Request Time	960	957	1041	1459	43074	43047
Request time increase %	N/A	0%	8%	52%	4387%	4384%

Note: The highest power cap level (110W) was found to slightly increase the average power use of the web server by 0.4W (0.4%), similar to results from LINPACK testing in the lab.

VALUE IN GRID BALANCING In addition to relieving strain on the grid caused by peak demand, fast-reacting DR is increasingly viewed as a tool to respond to the rapid variation of renewable energy sources, such as wind and solar photovoltaics (PV) on the grid. To provide the quick response required to react to these sources, markets have formed for ancillary grid services, such as frequency regulation, voltage support (providing reactive power), and reserve operating capacity (historically known as “spinning reserves”).

FAST RESPONSE TIME The inherent intelligence and communications abilities of servers offer the potential for quick, flexible, and complex DR behavior, with limited additional hardware costs. As power grids integrate more renewable resources, which are intermittent by their nature, increasingly quick and fast-acting DR capabilities will be needed. Data centers could be good resources for fast acting DR.

UPFRONT COST OF DCIM TOOLS ARE EXPENSIVE BUT PROVIDE MULTIPLE BENEFITS
Although DCIM tools are expensive, they provide many other business functions and features like asset management, etc. Power capping is a one of features of the software. Once DCIM tools are used in a data center, adding power capping features may not be very expensive.

LESSONS LEARNED

There are a few gaps for managing server power for DR at present that should be addressed before implementing DR software for data centers on a large scale:

Power system operator cannot be expected to decide how to limit each of its servers. Such information needs to be supplied by the data center operator, including the amount of power curtailment possible, and performance indications, such as response time, duration, and pricing signals. What's more, this cannot be expected using the tools that are currently available for local server management. The technology evaluated in this study elevates these tools to commands that can be called by software, but the operational knowledge to set power limits is still absent.

Additional testing is recommended. This is to evaluate power capping with additional real-world applications, such as e-mail server, database, etc., so the impact to workloads with different needs may be evaluated (comparing processor-intensive workloads to memory or data-limited applications). Such a study could be performed in both laboratory and field conditions, so that basic functionality, response to stochastic load, and user impacts can be quantified.

Development of more robust DR signal communication pathways. Future technology should allow an intelligent load to communicate its current state and make an informed decision about how to respond.

What We Concluded

SMALL LEVEL OF POWER CAPPING MOST PRACTICAL

Preliminary testing in the laboratory benchmarks demonstrates that power capping can successfully limit a server's instantaneous power. When applied to a loaded web server in the field, power capping was able to reduce average server power by up to 18% (18.2 W) under more heavily loaded testing. Yet these power reduction levels increased the time required to deliver HTTP requests by more than one order of magnitude, but increase in time was smaller with smaller level of power capping. Also, the tests were conducted with continuous web requests whereas in practical applications, the requests are random in nature. Thus, power capping may be practical for a smaller level of capping, say 5 to 10%, for short periods of time.

Though power capping may provide longer-term DR in some scenarios where workloads are flexible, it is likely that it will find most practical use for shorter-term durations, even in seconds and minutes. Power capping was demonstrated to successfully limit the instantaneous power of multiple servers at a moment's notice. This capability may provide value to the grid (for example, to respond to grid emergencies or to provide ancillary services, such as frequency or voltage regulation) with the very quick response required to compensate for variable generation sources on the grid (such as wind and solar).

These Findings are based on the reports "Data Center Demand Response via Server Power Management," which is available from the ETCC program website, https://www.etcc-ca.com/reports