RC25268 (W1111-050) November 8, 2011
Computer Science

IBM Research Report

Enabling Real-Time Data Center Energy Management

W. El-Essawy, A. P. Ferreira, J. C. Rubio, T. Keller, K. Rajamani, M. Ware
IBM Research Division
Austin Research Laboratory
11501 Burnet Road
Austin, TX 78758
USA

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

Enabling Real-Time Data Center Energy Management

W. El-Essawy, A. P. Ferreira, J. C. Rubio, T. Keller, K. Rajamani, M. Ware, M. Schappert*,
*IBM Research—Austin, *IBM Research—T.J.Watson*

Abstract— The most desirable data center energy monitoring and management capabilities required for energy efficiency are today not affordable, or even possible, for a variety of reasons. We enable these capabilities by two new pieces of technology: a low-cost, non-intrusive and retrofittable power monitoring sensor for power panel circuits and a semiautomatic server-to-circuit mapping system employing the new sensor. Applications include accurate charge-back to clients, server-to-circuit mapping for more efficient equipment placement, and branch circuit power capping allowing the safe oversubscription of power for more efficient utilization of power capacity, resulting in the deferral of major capital expenses.

Index Terms—AC energy measurement, data center, energy management, energy monitoring, power capping

IBM Research-Austin Laboratory, 11501 Burnet Rd, Austin, TX 78758, E-mail: wrelessa, apferrei, rubioj, tkeller, karthick, mware, schap1@us.ibm.com

1 Introduction

The explosive growth of data centers over the past few years, both in size and power density, has resulted in the power delivery infrastructure becoming a prime bottleneck to both efficient energy usage and data center utilization. Since data centers consume approximately 2% of USA electricity today, and grew 20% to 33% in the period between 2005 and 2010 [1], data center energy growth poses pressing environmental as well as economic challenges.

Power consumption is a major component of data center operational cost, while data center construction cost is proportional to peak power capacity. Limiting peak power consumption to fit within a company's existing data center's power capacity at different levels of the power distribution network (and without crippling performance) defers building a new data center, deferring tens to hundreds of millions of dollars of construction cost.

Organizations have become increasingly reliant on their IT infrastructure. It is common for a server to support dozens or even hundreds of customers, and downtime can result in significant financial losses. The power distribution infrastructure contributes to significant downtime risk. The failure of a single branch circuit (e.g., because of a circuit breaker tripping) may result in bringing down all servers, storage, and network devices connected to it if proper design is not implemented. In more drastic situations, a trickle effect can result in disastrous failure data center failure if the protection devices are not setup correctly [2], [3]. This is why higher redundancy, often rated in reliability "tier" ratings defined by the Uptime Institute, is usually adopted in enterprise data centers [4]. Unfortunately, our experience is that many data centers and machine rooms do not have accurate mappings of which equipment is fed by which circuit breaker, due to the expensive and cumbersome manual tracking required to maintain these mappings.

In order to properly handle the dynamic IT load growth, many data centers utilize a power allocation scheme to track the maximum power required by IT equipment at any time. The total power allocated for IT equipment has to be within the data center power capacity, which is the maximum that can be provisioned by its power distribution network. Overestimation of the power allocation results not only in poor utilization of the data center infrastructure, but it also leads to stranded power [5]. Stranded power is the amount of available power that cannot be used because it is allocated to IT load but never gets actually used. While wasteful, this is safer than underestimating allocated power, which increases the risk of overloading the power distribution network. Since the financial consequences of taking such risks may turn out to be drastic, this is usually a risk never taken. The result is poor data center utilization and its attendant costs.

Both charge back and peak power capping necessitate mapping IT equipment to branch circuits (the underfloor cables running from power panel circuit breakers to the racks) feeding them. This is currently done by a manual process that is expensive, cumbersome and error prone. We achieve semi-automatic branch circuit identification (BCID) using an algorithm that allows us to identify a many-to-many relationship between different IT equipment and branch circuits in power distribution units.

We present a system that achieves these goals. A power monitoring system measuring the actual current consumption at branch circuits is an essential component in our system, where the actual consumption measurements provide the basis for the charge back

policy and the power capping mechanism. In order to control the power consumption at different levels of the power distribution network, a power capping mechanism interacts with the IT equipment connected to the specified branch circuit(s), and picks the servers previously determined via the BCID algorithm to be fed by the branch circuit with specific equipment power caps to meet the power capping target.

In short, in addition to providing simple energy measurements, this system is fast enough and flexible enough to enable additional applications such as power allocation, power shifting and trending and stranded power reduction [5], [6], [7].

In the rest of this paper, we present some brief background, present the design goals which address the challenges introduced in this section, and summarize the hardware and software architecture of our system. In "Applications" we present the monitoring, branch circuit identification and branch circuit power capping capabilities of our system, before touching on future work and concluding.

2 Background

Power measurement for the data center power delivery infrastructure is available commercially from many vendors (Eaton, Schneider, Emerson Electric and others, with the Eaton Energy Management System Upgrade Kit [8] as one example). These power measurement systems are designed as tools for static power allocation and human interaction, so high speed is not a useful or provided function. The reporting interval of the current systems is 15 seconds or more. This reporting rate is too low for newer uses like dynamic power allocation or protection. These commercial systems are provided as a built-in feature in a power distribution unit or as add-ons. Current sensing is done using costly current transformers, contributing significant cost to these systems.

Methods for circuit identification for AC circuits have been in production for decades. Most solutions inject a high frequency signal into the power circuits. The detector has to identify the wires that contain the transmitted high frequency signal. This technique is not very robust because it is susceptible to interference and requires filtering otherwise more than one circuit is detected. A more recent method is load modulation, where the AC current is modulated by changing the load to generate an identifiable signature for the receiver. The detector uses some form of frequency detection to determine if the signal is present. One example of this approach to modulate the power consumption of a server is to change the CPU utilization. Microsoft's Red Pill [9] uses this approach, by inserting a software agent that periodically changes the CPU load. The major limitations of this approach lie in the need of IT access to the server, and when the server is heavily utilized, there is no available capacity to be modulated. Red Pill also suffers from the low reporting rate of the measurements that increases the detection time to hundreds of seconds.

3 Design Goals

This work began as part of a joint IBM and US Department of Energy project to improve data center energy efficiency (see [10], also the Acknowledgement) with a goal of producing a sensor system design that could be marketed at one-tenth the cost of systems available in the market. The motivation for this goal results from the scarceness of branch circuit level power measurements available for analysis today. While the need for better power distribution utilization (reducing stranded power [6], [11]) is well known [7] and schemes for improvement have been proposed [12], there is insufficient data to test these ideas. Discussions with data center facilities operators led us to the realization that the reason was the relatively high cost of power measurement systems for data center power distribution units (PDUs).

These discussions also led us to realize that the management schism existing in legacy data centers between the IT organizations and the Facilities organizations meant that retrofitting power monitoring and branch circuit identification (a Facilities function) had to be do-able by Facilities staff without IT staff involvement.

Our system design goals were meant to address cost and retrofit concerns, while providing equal or better accuracy than present-day systems. Existing solutions use expensive (on the order of $20-$30 each) sensors based on current transformers, in contrast to our sensors that use an order of magnitude less expensive off-the-shelf components. Multiple decisions were taken in order to reduce the system components: 1) A low cost, highly integrated microcontroller with a high sampling rate and accurate analog to digital conversion capabilities. The microcontroller also provides a USB interface to enable in the field firmware upgrades. 2) A high degree of multiplexing at the inputs was made to amortize the cost of digital and analog components over a larger number of measurements. 3) The processor board is powered through the USB, eliminating the need for extra power supplies. 4) An off-the-shelf computer board with computational capabilities high enough to multiplex multiple processor boards on one computer engine.

Another design goal of the system was that it be able to be installed in energized power distribution units since deenergizing a power panel is a major undertaking in most data centers. Working in live power panels is an inherently life threatening activity. The U.S. OSHA and other authorities require wearing protective gear during all work with energized PDUs. This protective gear varies according to the category of arc-flash risk assigned to the individual PDU. The gear ranges from protective gloves and minor clothing requirements at the minimum risk category up to very thick protective gloves and an all body protective suit and welder-style helmet. The sensors must be installable while wearing some version of this gear.

4 Power Monitoring Architecture

4.1 Hardware

The system consists of three components: a current sensor installed on a branch circuit in the PDU, a signal processing board (SPB) collecting and digitizing the signals, and a commodity, general purpose, small form factor computer (SFFC) running Linux that captures the data and provides a user interface.

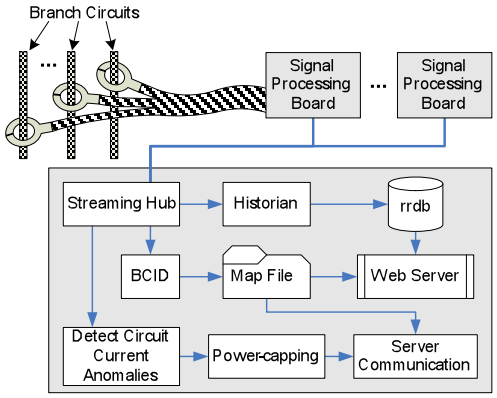

Fig. 1. Block Diagram of the System Architecture.

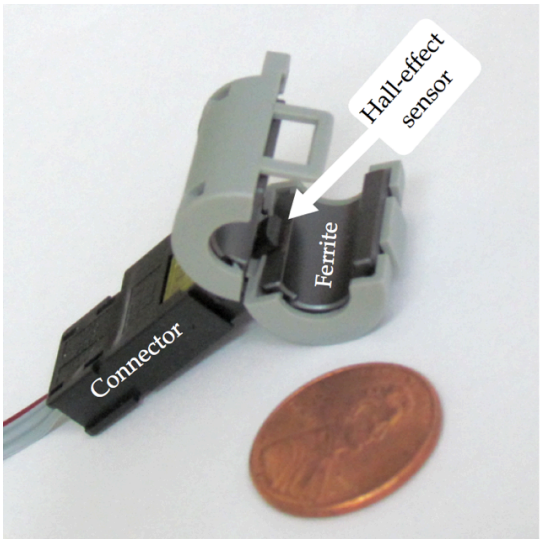

Fig. 2. The Power Sensor

Figure 1 is a schematic showing these as well as the software components of the SFFC. Shown in Figure 2 is the sensor combining an off-the-shelf clamp-on ferrite typically used in line cord noise filtering applications and a ratiometric Hall effect sensor, with a combined cost of $2.50 in small quantities. Using clamp-on ferrites enables non-intrusive deployment in the PDUs, as shown in Figure 3.

The SPB is a custom printed circuit board used to measure the instantaneous RMS (root mean squared) power consumed by branch circuits. It does so by measuring the current that flows through the cable using Hall effect sensors. This board is designed around a Texas Instrument MSP430 microcontroller. The SPB board also includes low-pass analog filters used to remove noise introduced by the environment. Multiple SPBs may be connected to the small form factor computer via USB communication ports.

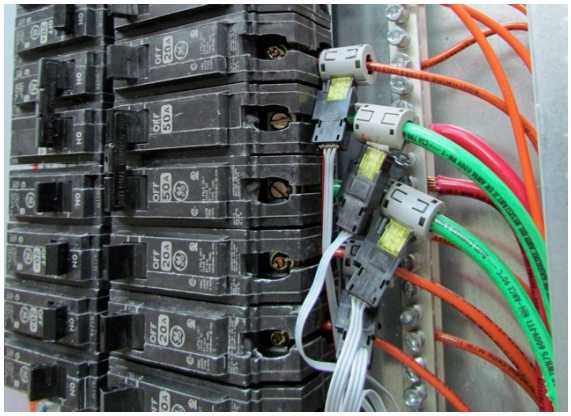

Fig. 3. Power Sensors Installed in Power Panel

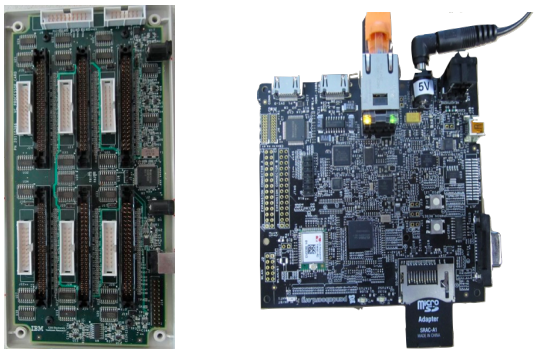

Fig. 4. The Signal Processing Board and Small Form Factor Computer

The components are installed inside, or in close proximity to the PDU, which typically contains multiple power panels. Each power panel has a (standard) capacity of 42 branch circuits. Current is sampled concurrently on all branch circuits in an interleaved way. Data is streamed from multiple SPBs, each monitoring up to 42 branch circuits, to a SFFC associated with the PDU. The accuracy of the sensor/SPB combination is +/- 5% above 3A with a resolution of .15A. Averaging multiple samples significantly improves recorded resolution. The SPB prototype and a Pandaboard SFFC are pictured in Figure 4.

4.2 Software

The SPB microcontroller uses 12 analog to digital conversion channels augmented with analog muxes, which allows the system to monitor the 42 branch circuits present in conventional power panels. During a 16.7 ms window (a 60 Hz AC cycle), the system samples every branch circuit channel 24 times. This corresponds to a sampling frequency of 1.44 kHz, which allows us to capture energy content up to the 12th harmonic frequency. At the end of the 16.7 ms interval, the system computes the cycle RMS using the expression

$$I_{rms} = \frac{1}{N}\sqrt{N\sum(I_i)^2 - \left(\sum I_i\right)^2}$$

where **N** is the number of samples in the cycle and $\mathbf{I}_i$ is the sample from the current sensor. Once the RMS values are computed, the packetized value is buffered for transmission to the SFCC via the USB link.

On the SFCC computer, a real time module receives the RMS packets, which are timestamped and forwarded to the multiple components that will consume them. As shown in Figure 1, the historian module generates statistics for the system and stores the values in a round-robin database (rrdb) [13] every second. The round robin database is configured to automatically aggregate current measurements to longer time intervals (e.g., seconds, minutes, hours, days, weeks, months).

The power monitoring and management applications that require access to historical data access the rrdb for all measured data. For applications that need access to real time measurement stream (at 60 Hz granularity), a streaming module intercepts the branch circuit current measurement streams, and forwards them to the specific application code. The SFCC also runs a web server to provide a user interface.

5 Applications

5.1 Monitoring and Protection

These components enable "managing the forest" by first providing accurate and fast measurement. Monitoring power at the branch circuit level means monitoring it at the rack level. Many data center racks house a single client. One effective way to reduce the energy consumption in a data center is to apply a fair charge-back policy for energy. Almost every data center today charges clients for energy usage by a simple formula in which square footage dominates; where the monthly charge is dominated by the space allocated by their servers and other IT equipment, irrespective of the load or the actual power consumed by their IT equipment. Lacking incentive to save energy results in unjustifiable waste of energy that can be eliminated by a fair charge back mechanism based on the actual energy consumed by each department. Branch circuit monitoring ties energy use to clients.

Tracking branch circuit-level power consumption over time enables the balancing of phases from power panels, detecting historically available power in circuits to avoid stranded power and guide new equipment placements. Further, it can also provide a critical input to datacenter planning for expansions.

Fast measurement also enables reduced power margins and corresponding reductions in stranded power in the following manner. Stranded power can be reduced by implementing a safe mechanism for eliminating power overdraws during server power supply failures. As one example, redundant power systems can result in stranding 50% of power capacity. If one (or even more) power supplies fail for a server, a redundant power supply suffices. However, in the facility where wasteful power margins are minimized by oversubscription, the additional load placed upon the redundant power supply can exceed the effective current capacity of the branch circuit feeding the redundant power supply. This situation can be prevented by means of dynamic power cap for the server. The system described later in the paper provides a power cap for one or more devices fed by every branch circuit.

Circuit breakers are used to preserve the integrity of the power distribution network by removing failures like short-circuits or overloads. When a circuit breaker trips, it disconnects the branch circuit potentially creating service disruptions by deactivating servers. Power allocation to the servers has to guarantee that in any foreseeable condition the total power consumed is higher than what is provided by the branch circuit (circuit breaker limit). The value of the peak power is usually much higher than the average power leading to large stranded power.

Our system is able, in case of overload, to dynamically change the peak power consumed by each server guaranteeing that the total power consumed is equal or below the available power. The detection of the overload condition, change in power cap and response by the server is fast enough to make the event undetectable by the circuit breaker. Interestingly, circuit breakers are designed to trip quickly in the event of a short but relatively slowly in the event of a current over load. Thus a window of time exists in which to draw down the over current to below the circuit breaker's limit. One typical circuit breaker can endure a 25% overload for a minimum of 200 seconds and a 100% overload for 35 seconds [14].

5.2 Branch Circuit Identification

Knowing the power distribution topology as it winds through the facility to the IT equipment as well as knowing the real-time load at each level of this infrastructure is necessary for optimizing a) management of available power, b) failure scenario responses, c) equipment placement and d) data center planning. Modern servers have power measurement capabilities but the real-time load within the facility power distribution infrastructure is unknown. This is because while instrumentation can be bought, it is too costly to enjoy widespread deployment. A further hindrance is that installing such instrumentation is disruptive to operations.

Our technology for doing this branch circuit identification (BCID) enables many power management applications including: load-aware power distribution, balancing power consumption across the three phases (for example by phase aware equipment deployment) and power capping at the branch circuit level. Also, while the power distribution topology is designed to be redundant at all levels, incorrect wiring can remove that redundancy and leave equipment vulnerable to being deenergized accidentally when one of the two redundant PDU's are deenergized, only to discover both "sides" of the equipment power supplies were connected to the deenergized PDU. BCID can be used to identify these miswirings and hence improve the system availability.

The manual identification of branch circuit maps is extremely difficult. In addition to being error prone, manually tracing a branch circuit underneath a raised floor for each branch circuit is frequently so labor intensive as to be impractical. (We demonstrate a technology that allows a power-varying signal generated by a server to be recognized at the SFFC.) The detection achieves server to branch circuit mapping. Similarly, a unique signal can be induced by an external source connected to a power outlet to establish the mapping between power outlets and branch circuits. This enables safe equipment installation in desired branch circuits, and optimizing electric phase load balancing, among other applications.

A practical BCID technology has to account for a set of requirements that comes from aspects of datacenters operations. One of those is the operational schism existing between the facilities areas (electrical, cooling) and the IT operations. Facilities personnel have little or no access to IT equipment or management systems. Patching in an intrusive technology like a software agent running in a server is also impractical, since it requires facilities staff to gain permission from IT staff for access to the system, as per [4]. It can also lower IT performance. Worse, some equipment, like network switches, cannot accept software agents. A good solution should be transparent to the hardware, OS and applications.

Our BCID technology leverages our power monitoring system by assuming that high speed sensitive power measurement is available at every branch circuit and employs a detection algorithm to construct a receiver that is sensitive enough to reliably detect a variation of 2.5W per branch circuit. At this power level, 2.5W, a simple USB device can be used as the power modulating device. This is key to achieving the requirement of transparency since our USB device can be hidden from the OS by only connecting to the power and ground signals of the USB connector. Figure 5 shows how BCID is integrated with the power measurement system.

Our BCID algorithm uses a synchronous detection [15] algorithm that measures the correlation of two signals; by assuming that the original transmitted signal is known, a much more sensitive and robust detection can be obtained than a more commonly used frequency detection algorithm (Goertzel, FFT, etc.). In our implementation the transmitter sends a copy of the signal used to modulate the power consumption to the SPBs via an external interface. The correlation is computed between the copy of the signal and the branch circuit measurements.

One beneficial property of synchronous detection is the very low run time cost (linear on the number of samples and on the number of channels). In the algorithm, a number of parameters can be optimized to manage false positives and false negatives. The integration time, proportional to the number of samples integrated to obtain a single output value, is a key parameter to determine sensitivity. Larger integration times improve sensitivity but decrease reaction time. False negatives are avoided by determining an integration time allowing a robust detection of 2.5W, which in our system is around 30 seconds.

Even though synchronous detection is very resilient to noise, it is not immune to it, so other techniques are applied to decrease the probability of false positives. We

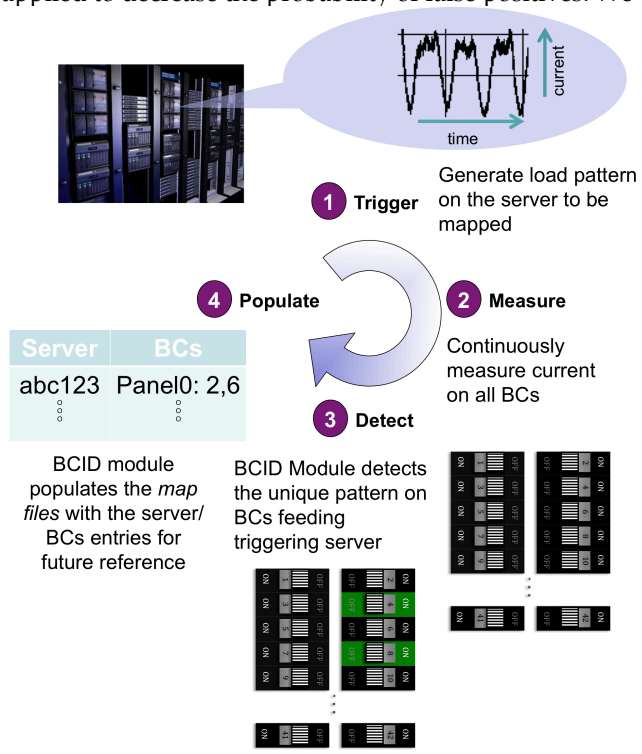

Fig. 5. Branch Circuit Identification Methodology

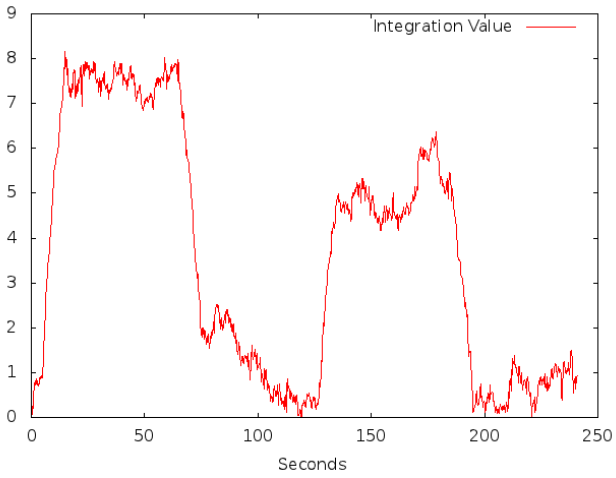

Fig. 6. BCID: Integration Value When Signal Present

determined that when a signal is present, the first and second derivatives of the integration value follow a very defined pattern. The second derivative has to be very small, meaning that the first derivative is constant over time or zero. A noise signal (a potential false positive) does not follow this pattern and presents large variations in the time derivatives of the signal. Figure 6 shows the sensitivity of the system in detecting two different signal

strengths when a 2.5W, first bump, and 1.875W BCID signal, second bump, is applied. Figure 7 shows another branch circuit where no BCID signal is applied (and also represents the worst case found in our experiments.) In almost all branch circuit tests, a simple threshold detector applied to the integration value is enough to discriminate. When augmented by a threshold detector on the second derivative, all false positives were eliminated in our experiments. The experiments used a USB device able to produce 5W of variation with two USB ports (2.5W per port) and the measurements were done on circuits carrying from 0.5A to 100A. The system under test had two redundant 208V power supplies with four branch circuits, so a 5W variation corresponds to .012A variation of the current in the branch circuits. Synchronous detection gives us the powerful advantage of detecting signals below the resolution of the sensing system.

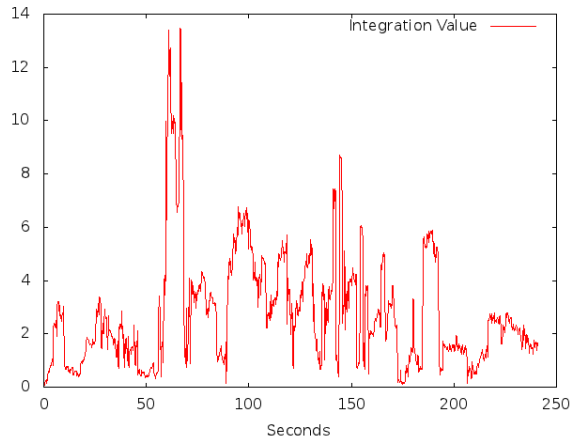

Fig. 7. BCID: Integration Value When No Signal Present

5.3 Power Management

In this section we provide a solution to improve data center utilization without suffering any deterioration in reliability. The way we achieve this is through utilizing branch circuit power capping, a mechanism that monitors the power consumption on a branch circuit, and guarantees that it doesn't exceed a certain preset level (power cap). While the term has been previously used for controlling the loads on compute servers [15], we extend this concept beyond the computer system level into the power distribution level.

Unlike the operation of a circuit breaker that trips when a preset (static or dynamic) value is exceeded, a system implementing branch circuit power capping controls the load connected to a branch circuit to keep levels within the branch circuit power cap. This way the power allocation scheme can be more aggressive and utilization can be improved without sacrificing reliability. We still utilize power protection mechanisms (e.g., circuit breakers) to protect the power distribution network.

The fact that we build the connectivity map for servers and the corresponding feeding branch circuits, power capping can be applied to the specific servers connected to the branch circuit suffering a power cap violation.

We utilize the power monitoring system to collect the branch circuits' power measurements. High frequency sampling of the power data, and availability of high-speed computation at the power panel level enable timely response to power cap violations.

As shown in Figure 1, the power-capping module continuously receives power measurement streams from the streaming module. When a specific branch circuit power exceeds the power cap (initially set for each branch circuit by the power allocation module by a facility manager), the power-capping module issues an alarm signal to the facilities user. The power-capping module reads the mapping file corresponding to the branch circuit, and looks up the list of servers connected to it, reading their power caps (initially set to their allocated powers). The power-capping module determines new server power caps. The power-capping module then uses the server communications module to send out the new server power caps. Each server then resets its power cap and the adjusted loads correct the branch circuit power to within preset levels.

We next describe an experiment that demonstrates using branch circuit power capping to protect the branch circuit from overcurrent in the case of a power supply failure. We apply a relaxed availability branch circuit power capping policy with a DVFS server power capping policy (both defined in Future Work).

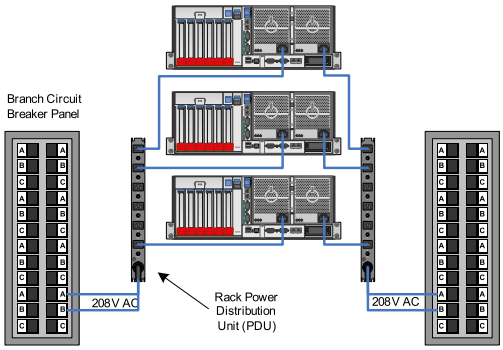

Fig. 8. Schematic of Test Machine Power Cabling

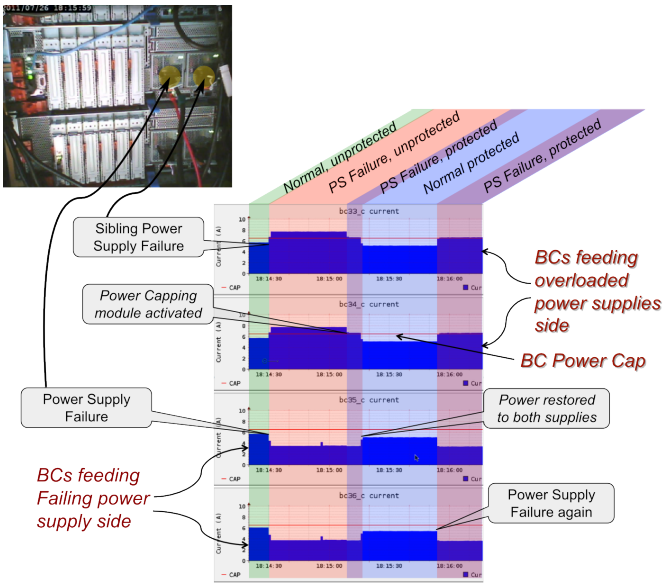

Fig. 9. Power Capping Experiment

The setup of the experiment is shown in Figure 8, where three IBM midrange System p POWER7 servers are powered by a redundant power distribution system. The IBM systems have redundant power supplies, each with separate 208 V feeds through rackmounted PDUs. The PDUs are each fed with a pair of branch circuits connected to a different power panel. Since all the servers in this setup are connected to the same power strips, a total of four branch circuits power all the servers.

The experiment starts after the BCID algorithm is applied, identifying the three "servers' connectivity," and generating map files fully populated for all feeding branch circuits. We emulate a power supply failure by unplugging the power cord connected to one of the three servers. Figure 9 presents the current measurements on the four branch circuits feeding the three servers. The two lower graphs present the current consumption feeding the side of the failing power supply (we call it Side A). The upper two charts present the current on the branch circuits feeding the other side (Side B). As shown in Figure 9, the power supply failure results in increased current flow on Side B to compensate for the power lost due to unplugging the power supply (and hence reduced measured current) on Side A. We model a power cap in this experiment by a maximum current limit, shown in this figure by a horizontal red line.

The experiment is first run without enabling the power capping module shown in Figure 1. Figure 9 shows the power supply failure causes a branch circuit power cap violation as the currents on Side B exceed their cap. Next, the power cap violation occurs while the power capping module is enabled. The module realizes that the side B branch circuits suffer a power cap violation, so it looks up the map files listing the connectivity of side B branch circuits, and identifies the servers supplied by them. The two branch circuits feed the same three servers, the map files are identical, and they both list the names of the three servers. The server processor on the machine with failed power supply contacts the server communication module to report the incident. The power capping module identifies the server which suffered the power supply failure. The power capping module determines a new server cap value for that server, and sets it by sending the value to its service processor through the server communication module. Within a second, the service processor on this server applies the appropriate voltage and frequency, and lowers the server power via its DVFS control. This in turn brings the branch circuit power to safe limits within the identified power cap.

We then plug in the power again to the power supply, power levels are once again evenly spread among all branch circuits, and the increase in current is maintained within safe limits.

6 Future Work

One area to investigate surrounds the policies for branch circuit power capping and server power capping and their interaction. The power capping policy is responsible to setting the server power cap values to meet the branch circuit power cap, as defined by the facilities personnel. This policy determines the victim servers to be throttled, and the degree of throttling relaxation. The new capability of being able to respond to current overages very quickly opens up a new class of *relaxed* policies in contrast to today's *strict* policies.

In strict policies, individual server power caps are statically set such that the sum of all server power caps is equal to the branch circuit power cap. This guarantees that no branch circuit power cap violation will occur, since the server power capping mechanism maintains load levels below their corresponding caps. While these policies are simple, they lead to either performance or stranded power problems.

Relaxed policies allow the sum of the server power caps to exceed the branch circuit power cap. The server power cap values can be set to accommodate a variety of considerations including different service level agreements across servers and not just for branch circuit-level capping. Consequently, momentary violations of branch circuit power caps can occur.. When such a violation occurs, the quick response of our infrastructure can trigger a new (potentially temporary) set of server power caps to return total branch circuit load to permitted levels in a safe manner (before overloading the branch circuit breaker). This way, server power caps can be leveraged for more functions (e.g. differentiated service) in addition to equipment protection, also allowing for higher levels of performance, and less stranded power.

Each of the following branch circuit power capping policies may be implemented as either a strict or a relaxed policy. *Fair throttling* sets the same power cap to all servers. *Priority throttling* applies server power caps according to a specific (static or dynamic) priority scheme, where lower priority servers suffer a lower power cap and performance. *Availability throttling* is a variant of the priority throttling policy, where the priority of a server with a failing power supply drops to the least level, in order to reduce the load on the surviving power supply.

Performance throttling sets the priority of servers connected to the branch circuit proportional to their performance per watt. *Proportional throttling* sets a power cap proportional to the server's rated power.

We see two classes of mechanisms that the server applies to keep its power within a specific server power cap value. *DVFS* sets the voltage and the frequency of the server to meet its power cap. In a simple implementation, the voltage and frequency will be set to a predetermined value that guarantees that the power never exceeds the server power cap. In a more sophisticated implementation, the voltage and frequency will be dynamically adjusted to optimize performance within the allowed server cap. In a *scheduling* mechanism a workload scheduler will distribute the load in such a way to keep the load within the specified server cap.

Further, we would like to investigate how dynamic policies could be employed in data centers to make them load adjusting components in a "smart grid" electric system.

7 Conclusion

We have introduced a new data center power monitoring and management framework enabled by new technology: a highly accurate and responsive branch circuit power sensing system coupled with a branch circuit identification methodology that enables a practical server-to-circuit mapping system for the first time. As Liu states "Power provisioning, power capping, and power tracking all depend on accurately accounting which server consumes power from which circuit." [9] Further, the responsiveness of the system enables the safe adoption of new relaxed policies that oversubscribing branch circuit power [6] [7] that should substantially reduce power allocated that is currently wasted and results in unnecessarily accelerating the construction of new data centers.

Acknowledgment

The work on developing the power monitoring sensor is funded in part by the Industrial Technologies Program of the DOE Office for Energy Efficiency and Renewable Energy under project number 10E0002897, "A Measurement-Management Technology for Improving Energy Efficiency in Data Centers and Telecommunication Facilities," H. F. Hamann, Principal Investigator. Dr. Hamann originated the idea of using a Hall effect sensor for branch circuit power monitoring.

References

[1] J. Koomey. "Growth in Data Center Electricity Use 2005 to 2010," Oakland, CA. Analytics Press, August 2011, http://www.analyticspress.com/datacenters.html

[2] F. Bodi, "'Super Models' in Mission Critical Facilities," Telecommunications Energy Conference (INTELEC), 32nd International , vol., no., pp.1-7, 6-10 June 2010.

[3] F. Bodi, "Large-scale Modeling of Critical Telecommunications Facilities and Data Centers," Telecommunications Energy Conference, 2008. INTELEC 2008. IEEE 30th International , vol., no., pp.1-8, 14-18 Sept. 2008

[4] S. Pelley, D. Meisner, P. Zandevakili, T. F. Wenisch, J. Uderwood, "Power Routing: Dynamic Power Provisioning in the Data Center," ASPLOS '10, Proc. ACM conf. on Architectural support for programming languages and operating systems.

[5] K. Rajamani, C. Lefurgy, S. Ghiasi, J. Rubio, H. Hanson, and T. W. Keller, "Power Management Solutions for Computer Systems and Datacenters," in Int. Symp. On Low-Power Electronics and Design, 2008.

[6] X. Fan, W. D. Weber, and L. A. Barroso, "Power Provisioning for a Warehouse-sized Computer," ISCA 2007, in Proc. 34Th Int. Symp. on Computer Architecture.

[7] C. LeFurgy, X. Wang and M. Ware, "Power Capping: A Prelude to Power Shifting," Cluster Computing 11, 2 (2008), 183–195.

[8] Eaton Energy Management System Upgrade Kit Product Literature, http://powerquality.eaton.com/Products-services/power-distribution/ems-info.asp?CX=5, 2008

[9] J. Liu, "Automatic Server to Circuit Mapping with The Red Pills," HotPower '10, 2010 Workshop on Power Aware Computing and Systems.

[10] H. F. Hamann, T. van Kessel, M. Iyengar, J. Y. Chung, W. Hirt, M. Schappert, A. Claassen, J. Cook, W. Min, Y. Amemiya, V. López, "Uncovering Energy Efficiency Opportunities in Data Centers," IBM J. Res. & Dev. 53, 19, 2009.

[11] S. Govindan, J. Choi, B. Urgaonkar, A. Sivasubramaniam, and A. Baldini. "Statistical Profiling-based Techniques for Effective Power Provisioning in Data Centers," EuroSys '09, In Proc. of the 4th ACM European conf. on Computer systems.

[12] X. Wang, M. Chen, C. Lefurgy, T. W. Keller, "SHIP: Scalable Hierarchical Power Control for Large-Scale Data Centers," PACT 2009, on pp. 91-100 of Proc. Int. Conf. on Parallel Architectures and Compilation Techniques, 2009 aton Energy Management System Upgrade Kit Product Literature, http://powerquality.eaton.com/Products-services/power-distribution/ems-info.asp?CX=5, 2008

[13] "RRDtool", http://oss.oetiker.ch/rrdtool/

[14] "Typical Circuit Breaker Trip Curve," Tyco Thermal Controls,http://accontrols.com/Typical%20Circuit%20Breaker%20Trip%20Curve.pdf, 2003

[15] M. L Meade, "Lock-in Amplifiers: Principles and Applications," IEEE Electrical Measurement Series 1 (Peregrinus, Stevenage, Herts, England, 1983.