Intel® Cloud Builders Guide
Intel® Xeon® Processor-based Servers
Service-Aware Energy Management in Cloud-Oriented Telecommunications Services Infrastructure

Intel® Cloud Builders Guide to Cloud Design and Deployment on Intel® Platforms

Service-Aware Energy Management in Cloud-Oriented Telecommunications Services Infrastructure

Intel® Xeon® Processor 5500 Series

Intel® Xeon® Processor 5600 Series

Audience and Purpose

This reference architecture outlines the usage of energy management technologies as part of planning, provisioning, and optimizing strategies in cloud data centers to reduce energy costs and to address carbon emissions for green IT goals. This reference architecture is direct result of collaboration with Intel Corporation and Telefónica Investigación y Desarrollo (Telefónica I+D). It is intended for data center administrators and enterprise IT professionals who seek energy management solutions to achieve better energy efficiency and power capacity utilization within new or existing data centers. The techniques and results described in this reference architecture can be used as an example to understand energy management solutions implemented with the use of hardware and software components. The reader should be able to develop appropriate energy management solutions based on the design options presented using the state-of-the-art Energy Management Solution that incorporates the concept of service awareness, with reliance on service performance measurements with Intel® Xeon® processor 5600 series-based servers implementing Intel® power management technologies.

Table of Contents

Executive Summary

Cloud computing is revolutionizing the way telecom operators manage their service infrastructure, moving from an isolated and vertical approach to an open and horizontal model. As a result of this transition, the infrastructure management features available to telecom operators have notably improved, and energy management is not an exception. However, despite the relevance of energy efficiency to telecom operators and the readiness of these technologies, they have been only timidly adopted so far due to their unpredictable effect on service performance. Fortunately, the introduction of service-aware energy management techniques is reversing this trend, allowing operators to implement energy saving tools while assuring service quality.

This reference architecture delves into the considerations by telecom operators driving energy conservation solutions in their service infrastructure with Intel® Xeon® processor 5600 series, with descriptions of typical target application scenarios. It also discusses how these scenarios may be addressed through the application of dynamic infrastructure management technologies, making use of the energy management features available in cloud-oriented platforms. This configuration regulates power consumption with Intel® Intelligent Power Node Manager based on workload demands.

The Intel Xeon processor 5600 series provides hardware and firmware based features, to implement lower energy consumption, which provides TCO benefits and realizes a smaller carbon footprint. Finally, this collaboration introduces a set of trial experiences, showing how direct energy savings of up to 27 percent may be attained on different telecommunications services with minimal and controlled performance impact.

Introduction

Cloud computing is arguably one of the hottest topics in the telecommunications industry nowadays. From the service perspective, telecom operators are refining their traditional hosting offerings with a cloud computing twist, complementing them with their data communications services to compose end-to-end solutions. Moreover, operators have also realized the business and operational advantages derived from cloud computing, and have started transforming their service infrastructures towards this model.

Besides the ordinary benefits derived from cloud computing adoption, this evolution has also entailed a significant paradigm shift in the way telecommunications services are deployed, moving from the traditional silo approach towards horizontal and open platforms. By opening up their service infrastructure, operators have gained greater platform access, since the management features of the different infrastructure elements and layers are now exposed.

Energy management is one of these newly accessible administration capabilities, since traditional telecom systems are typically capable of implementing standalone energy conservation practices only. However, despite the commitment of telecom operators to energy efficiency, state-of-the-art energy management solutions incorporate the concept of service awareness, with reliance on service performance measurements to trigger the application of energy saving measures. The goal of this paper is precisely to present these advanced energy management solutions found in cloud-oriented platforms with the following usage models:

1. **Perform real-time server energy usage monitoring, reporting, and analysis** to get continuous and actual energy usage visibility via monitoring of the servers via Intel Data Center Manager along with instrumentations resident on the server for Intelligent Intel Node Manager. The power reporting capabilities allow analysis and enable logging/tracking of power energy for cost and carbon emissions reductions decisions.
2. **Power guard rail and optimization of rack density** by imposing power guard to prevent server power consumption from straying beyond a preset limit. The deterministic power limit and guaranteed server power consumption ceiling helps maximize server count per rack and therefore return of investment of capital expenditure per available rack power when the rack is under power budget with negligible or no per server performance impact.
3. **Static and dynamic power capping** by applying significantly lower power caps to lower power consumption and heat generation when unforeseen circumstances like a power outage on the utility mains or a cooling system failure occurs. In these scenarios it may be appropriate to set aggressively lower power caps though performance would be affected. The use case illustrates how it works at a data center location or a group of servers for business continuity decisions for critical conditions.
4. **Power optimized workloads** by taking into account service priority SLAs when applying power caps, it becomes possible to prioritize services as desired, favoring them up to the maximum desired degradation in standard service performance.

The approach is to match actual performance against service level requirements.

5. **Data center energy reduction** through Power Aware Support for Multiple Service Classes showcases the ability to enforce multiple SLAs across different populations of users with different priority workloads. Workloads that ran over a period of eight hours realized 27 percent less energy consumption.

For this purpose, this paper presents the main application scenarios for energy management solutions in a telecom operator's infrastructure, and details the power monitoring and control tools available on cloud-oriented platforms, showing how they may be combined to achieve a holistic service-aware management solution. Next, it presents the reference implementation and trial experiences carried out by Telefónica I+D in cooperation with Intel with a goal to validate these tools and assess their impact on energy savings and service performance.

Energy Management Scenarios for a Telecommunications Operator

Despite the efforts of telecommunications equipment vendors in reducing energy consumption for each new generation of systems, improvements in the energy efficiency of telecommunications infrastructure have not kept up with the increase in the number of subscribers and associated demand for resources. For example, a European telecommunications company reports that, although the electricity use per information unit decreased between 2003 and 2005 by 39 percent per year, such effect was more than negated by an increase in bandwidth requirements of 50 percent annually.

The net effect is that the overall energy consumption of telecom operators keeps steadily increasing, and such growth incurs significant business, operational, environmental, and regulatory issues that adversely affect telecommunications facilities. Some of the most relevant issues or energy efficiency motivations for telecommunications operators are the following:

Profit margin improvement: Due to the rise in the price of energy sources, the increase in demand, and the bigger share of renewable energy (costlier to produce), energy cost has not only experienced a notable increase during the last few years, but this trend is also expected to continue in the near future. Moreover, due to this increase in the price of energy, energy costs start surpassing equipment costs if we consider the power consumption during the whole lifespan of a system.

- **Extending the lifespan of current telecommunications sites:** Telecommunications infrastructure has experienced a significant evolution towards smaller form factors and greater density rates, thus multiplying by five the power demand per unit of area during the last 10 years. In addition, electric companies are expected to start imposing power caps to such facilities due to limitations in their grids. Consequently, the adoption of energy efficiency improvement initiatives, allowing the growth of infrastructure capacity and performance levels while maintaining or even reducing power consumption, becomes a key element for making use of existing sites and extending their lifespan.
- **Sustainable development and corporate social responsibility:** Telecommunications are perceived as a major tool in the fight against climate change, since they present an enormous potential to reduce the Greenhouse Gas (GHG) emissions of other industrial sectors by optimizing production and logistics and avoiding the transportation of people and goods. Therefore, providing these services as efficiently as possible by optimizing the energy consumption of telecommunications infrastructure itself is a key factor to maximize this positive contribution.

Compliance with current regulations and preparation for future policies: The European Union and its member states have adopted action plans against climate change based on the promotion of renewable energy sources and the reduction of energy consumption within the 2020 timeframe. Therefore, the adoption of energy efficiency improvement plans enables operators to take advantage of the current incentives for power saving initiatives, while preparing their infrastructures for prospective restrictions in future regulatory environments.

- **Differentiation and product positioning:** According to recent surveys, 26 percent of European online users would be happy to pay more for green products, or would be influenced by a company's environmental policy during their purchase decision-making process. In the corporate sector, surveys show that environmental awareness has also caught on in European companies. A strong willingness to pay a premium for green IT suppliers is found across companies of all sizes and sectors, with 52 percent of the surveyed enterprises willing to spend extra. In general, customers would prefer an ecologically responsible product. Therefore, operators must adapt to take their customers' demands into account and take advantage of these new business opportunities.

As the telecommunications infrastructure grows to meet user demand, the associated consumption of energy is growing as well and poses serious

operational and business issues. The energy-aware infrastructure management approach proposed and evaluated in this project intends to serve as a tool to address these topics, adapting power consumption to service demand and maximizing energy savings.

Architecture

Test Bed Overview

Figure 1 depicts the test bed infrastructure used in the use case evaluation tests described in this section. As shown in the diagram, we may distinguish between two main system groups:

- The service infrastructure itself based on the four Intel® Xeon® processor 5500 series based servers (iBOI1 to iBOI4) equipped with Intel Intelligent Power Node Manager technology provided by Intel Corporation for this proof of concept.
- Two auxiliary systems (HPM and iBOIx), where the management systems and simulators reside.

Additionally, a separate Network Attached Storage (NAS) volume for virtual machine storage completes the testing infrastructure.

Service Infrastructure

As shown in Figure 1, the four Intel Xeon processor 5500 based servers supporting the service infrastructure have been divided in two identical groups:

- A "Gold" group simulating the infrastructure for high-priority services composed of the iBOI1 and iBOI2 servers.
- A "Silver" group simulating the infrastructure for standard services composed of the iBOI3 and iBOI4 servers.

Additionally, these two server groups have been put together into a higher-level group, named "Green", which represents the whole service facility.

Each of these servers has been installed with VMware ESX 4.0*, and two separate server clusters have been created at virtualization level:

- A "Gold" cluster simulating the virtualized infrastructure for high-priority services, composed of ESX hosts iBOI1 and iBOI2.
- A "Silver" cluster simulating the virtualized infrastructure for standard services, composed of ESX hosts iBOI3 and iBOI4.

A sample service instance has been deployed on each of these virtualized server clusters, simulating a high-priority and a standard workload respectively. The Telefónica Rate Profile application for premium mobile content (APC, for Aplicación de Perfiles de Cobro) was chosen as the sample workload. Each APC instance is composed of 4 front-end nodes in a high-availability cluster configuration and a message queuing node for intra-cluster communication. Additionally, the Gold group includes an enterprise database node. Each of these APC nodes is an independent Solaris 10* x86 virtual machine.

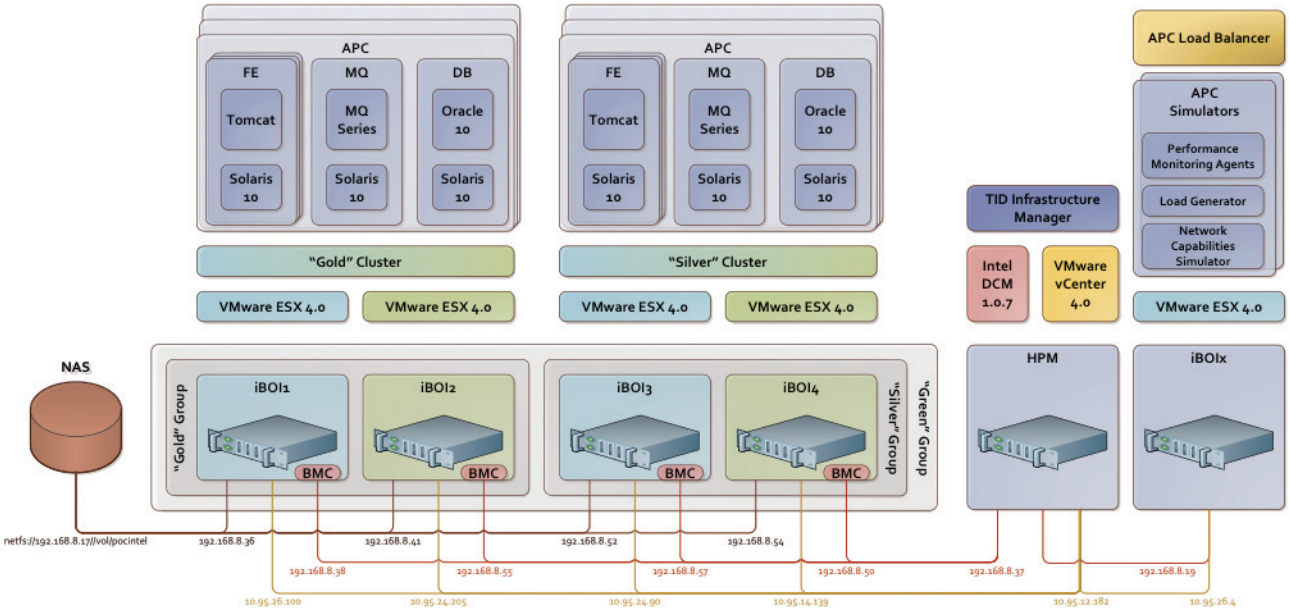

Figure 1: Physical Architecture of Test Bed Setup

Management and Auxiliary Systems

Management Server

The infrastructure management platform used in the proof of concept has been installed in an independent server (HPM) running Windows Server 2003* SP2. As shown in Figure 1, the software installed in such server includes:

- Intel® Data Center Manager (Intel® DCM) v1.0.7
- VMware vCenter* v4.0.0
- TID's Dynamic Infrastructure Manager. For the tests conducted for this paper, this software was installed on a VMware Virtual Machine with 4 CPUs, 6GB RAM and 50GB hard disk space. Windows 2008 R2* 64-bit was the operating system chosen.

APC Simulators

The APC Simulators are a suite of validation and testing tools for the APC service. They comprise the load injectors for generating signaling traffic, the emulators that simulate the core network capacities with which the APC service communicates, and the Performance Monitoring Agents with which TID's Dynamic Infrastructure Manager interacts to obtain the KPI information required to take management decisions.

As shown Figure 1, two APC Simulators (apc_sim1 and apc_sim2) have been deployed in the test bed, each installed in an independent Solaris 10 x86 virtual machine. The apc_sim1 simulator injects traffic to the high-priority (Gold) APC instance, whereas the apc_sim2 simulator feeds the standard-priority (Silver) instance. Both VMs run in the iBOIx server, which has been virtualized with VMware ESX 4.0.

Load Balancer

All traffic injected by the simulators to the APC front end service layer is balanced by Pound7, an open-source application that can behave as HTTP reverse proxy or as load balancer. Pound has been installed and configured in an independent VM based in a Debian GNU/Linux* OS, and deployed in the iBOIx server together with the APC simulators.

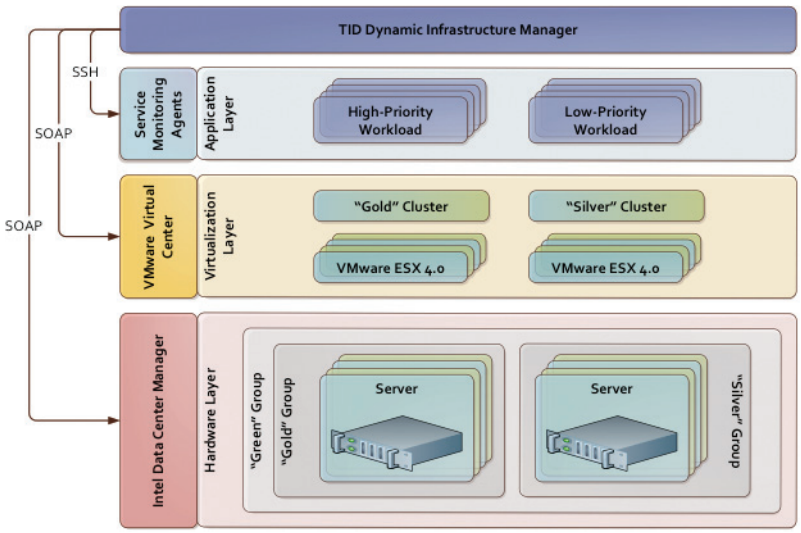

Figure 2: Test Bed Power Management Architecture

The load balancer distributes the overall signaling load produced by the simulators across the active APC front-end instances. Moreover, it periodically monitors the availability of the different front-end nodes, routing traffic to them when they are available and removing them as traffic recipients as soon as it detects that they have been suspended by the Dynamic Infrastructure Manager.

Service-Aware Energy Management Overview

Figure 2 summarizes the logical architecture of the energy efficiency optimization solution evaluated in this proof of concept.

As shown in Figure 2, the Dynamic Infrastructure Manager is in charge of enforcing the energy management policy. The manager interacts with the systems under control at three different levels:

- **Hardware level,** including real-time power consumption monitoring, the management of the system power states, and any additional energy-saving features implemented by the computing elements (e.g. power capping). As shown in the figure, these monitoring and control capabilities are based on the features offered by Intel's power management technologies.
- **Virtualization level,** including the monitoring of guest-level resource utilization and the management of virtual system placement, reconfiguration, and the allocation of physical resources to virtual systems. As shown in the figure, these monitoring and control capabilities are based on the features offered by VMware vSphere 4*.
- **Application level,** including aspects such as monitoring of service-specific KPIs and managing application scalability and workload placement. As shown in the figure, these monitoring

and control capabilities are based on the features offered by Telefónica I+D's service monitoring agents. Sections 5, 6, and 7 detail the energy management mechanisms available on each level.

This architecture reflects the two main power management use cases implemented in this proof of concept:

1. Dynamic power capping in the form of a sub-group time varying power allocation from a global power target with QoS driven prioritization.
2. Power aware dynamic reconfiguration with server parking, where servers not needed to meet application SLA are placed in hibernation.

The middle layer carries out hibernation control through the VMware vCenter DRS/DPM API. This was handy given that VMware was already in use supporting the virtualization layer.

The lower [hardware] layer implements dynamic power capping. Normally power capping is challenging to implement. We took advantage of built-in policy supported by the Intel DCM SDK to reserve a portion of a global power quota to two sub-pools out of a global pool, and hence this capability was trivially easy to implement. Each sub-pool represents a class of service, labeled "Gold" for premium service and "Silver" for best effort. Left alone, the Silver sub-pool consumes more power because of higher workload demand and because certain bookkeeping tasks are also run in these machines. This is not a problem during low demand. However, during peak usage the Gold QoS may deteriorate below the SLA. Intel DCM has a feature to impose power caps by sub-group. We used this feature to prevent the Silver sub-pool from starving the Gold group.

Intel® Power Management Technologies

In this section we present the power monitoring and control features offered by the combination of Intel Intelligent Power Node Manager and Intel DCM. For further information on these elements please refer to Appendix A.

Intel® Intelligent Power Node Manager

Servers based on the Intel Xeon processor 5500 series carry a capability to perform real-time power consumption monitoring and control, thereby enabling data center operators to trade off power consumption level against performance. This capability is valuable from two perspectives. First, power capping brings predictable power consumption within the specified power capping range, and second, servers implementing power capping offer actual power readouts, as a bonus – their power supplies are PMBus*-enabled and their historical power consumption can be retrieved through standard APIs.

Intel® Data Center Manager

Integration with the Intel DCM SDK, allows dialing the power to be consumed by groups of over a thousand servers, allowing a power control authority of tens of thousands of watts in data centers.

How does power capping work? The foundation for power management resides in CPU voltage and frequency scaling, or DVFS and is implemented by the Intel Xeon processor 5500 series architecture.

More likely than not, CPUs represent the most energetic components in a server. If we can regulate the power consumed by the CPUs we can have an appreciable effect on the power consumed by the server as a whole. Multiply this control over the thousands of servers in a data center. Through this mechanism, we can alter the power consumed in that data center in significant ways.

Intel® Data Center Manager Interface

Intel DCM exports a Web Service interface that enables external management systems, like the one evaluated in this proof of concept, to make use of it for power and thermal server management.

In the context of the evaluation activities undertaken in this project, the Java* language stubs for that Web Service interface were obtained from its WSDL definition by means of the Apache* Axis2Java tool. A group of Java libraries was built with these stubs, the Intel DCM Wrapper libraries, ready to be used and integrated with third party software, allowing easy operation and monitoring of the hardware elements managed by an Intel DCM instance.

Specifically, the Intel DCM Wrapper libraries implement the methods that allow executing the following actions:

- **Connect** with Intel DCM.
- **Find** an entity managed by Intel DCM.
- **Get** an entity's object properties.
- **Get** an entity's policies.
- **Update** or **delete** an entity's policies.
- **Get** the power consumption, temperature, etc. history of an entity for a certain period.

The Intel DCM Wrapper libraries allowed easy integration of the Dynamic Infrastructure Manager (DIM) with Intel DCM. Furthermore, an additional Java-based tool named Power Consumption Logger (DCMLogger) has been developed in order to register the average power consumption of an entity managed by Intel DCM during a certain period of time. The power consumption records obtained through that tool have been used in the use case evaluation activities.

Virtualization Level Power Management

In this section we present the power management features implemented by VMware vSphere 4: the operating system power management (OSPM) capabilities implemented by the VMware ESX 4 hypervisor and the VMware Distributed Power Management solution that VMware vSphere 4 offers for virtualized server clusters.

Hypervisor-Driven Power Management

To improve CPU power efficiency, VMware ESX hosts can be configured to dynamically switch CPU frequencies based on workload demand. If this feature is activated, as shown in Figure 3, the VMware ESX hypervisor makes use of the processor performance states (P-states) made available to the VMkernel through the ACPI interface to carry out dynamic processor voltage and frequency scaling.

Figure 3 shows the OSPM activation and configuration screen, available on the VMware vCenter management console under the advanced software settings for host configuration. By setting the Power.CPUPolicy property to dynamic, the VMkernel will start optimizing each CPU's frequency to match demand in order to improve power efficiency but not affect performance. When the CPU demand increases, this policy setting ensures that CPU frequencies also increase.

Two additional settings may be controlled. The Power.MaxCpuLoad property allows setting the CPU utilization threshold under which DVFS will be applied. Whenever CPU utilization surpasses that threshold, the processor will always operate at maximum frequency. The Power.TimerHz property allows configuring the CPU utilization polling interval. Higher timer frequencies will result in a finer-grained monitoring interval.

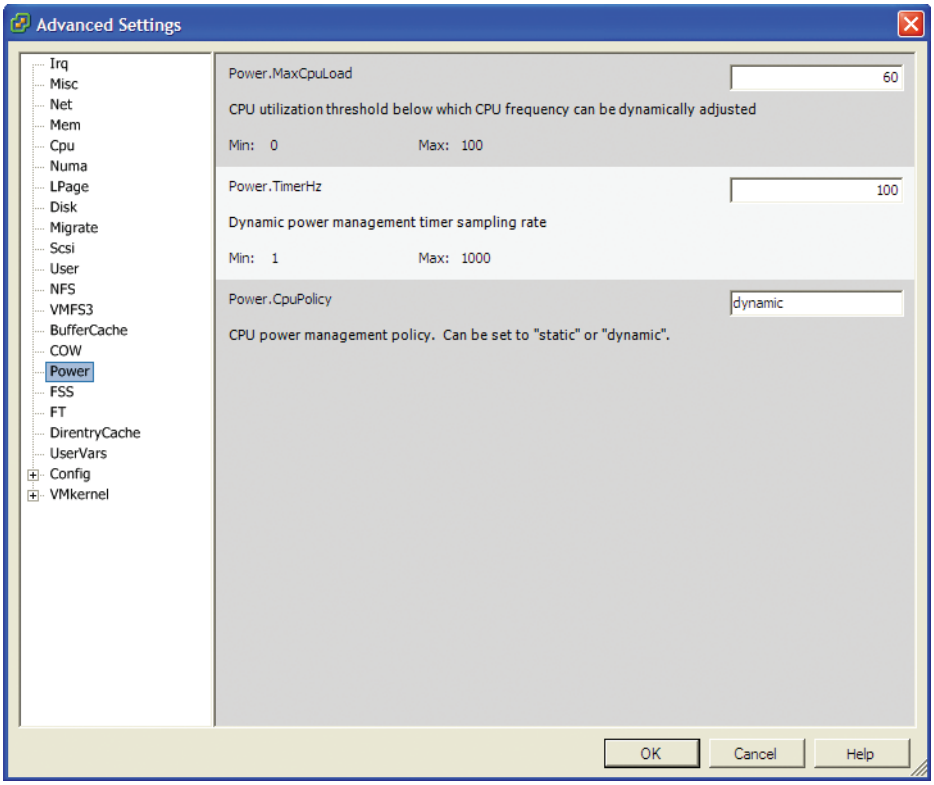

Figure 3: Test Bed Power Management Architecture

Distributed Power Management

Starting with version 3.5, the VMware Distributed Resource Scheduling (DRS) tool set includes an energy saving feature called Distributed Power Management (DPM)7. As shown in Figure 4, this tool is based on the virtual machine migration and deployment optimization features implemented by DRS. The difference is that, in this case, instead of dispersing virtual machines to optimize cluster resource availability, DPM dynamically consolidates virtual machines in the minimum possible number of ESX servers required to supply the demanded resources, powering off idle cluster elements. When resource demand increases, DPM activates the required number of dormant servers through IPMI or Wake-on-LAN, migrating virtual machines to these hosts as soon as they become available.

The VMware DPM behavior is governed by a series of heuristics, whose objective is to maintain the utilization level of active cluster members within a certain range (between 45 and 81 percent utilization, by default) by shutting down or activating VMware ESX hosts. Utilization is calculated as the quotient between resource demands (CPU and memory) and the capacity available on each node. The demand for resources does not only consider the actual resource usage of the virtual machines running on the hosts, but also an estimation of resource demands that are not being granted due to lack of available resources.

By default, DPM analyzes resource utilization every five minutes (DRS polling interval) and issues management actions or recommendations (depending on the chosen DPM automation level) if necessary. It should be noted that, to avoid compulsive virtual machine migration and node power-off and reactivation in case of fast variations in infrastructure load, DPM does not evaluate instant utilization, but analyzes the average values during a certain interval. Node shutdown is much more conservative than node activation. By default, the average utilization value during the last 40 minutes is analyzed for taking node power-off decisions, whereas just five minutes of history are considered for node power-on. Once the necessity to increase or decrease cluster capacity has been detected, DPM makes use of DRS in "what if" mode to analyze all the possible configurations, recommending or applying the most efficient one in terms of the number of changes (virtual machines to be migrated and nodes to be started or powered-off) required to achieve the target utilization range.

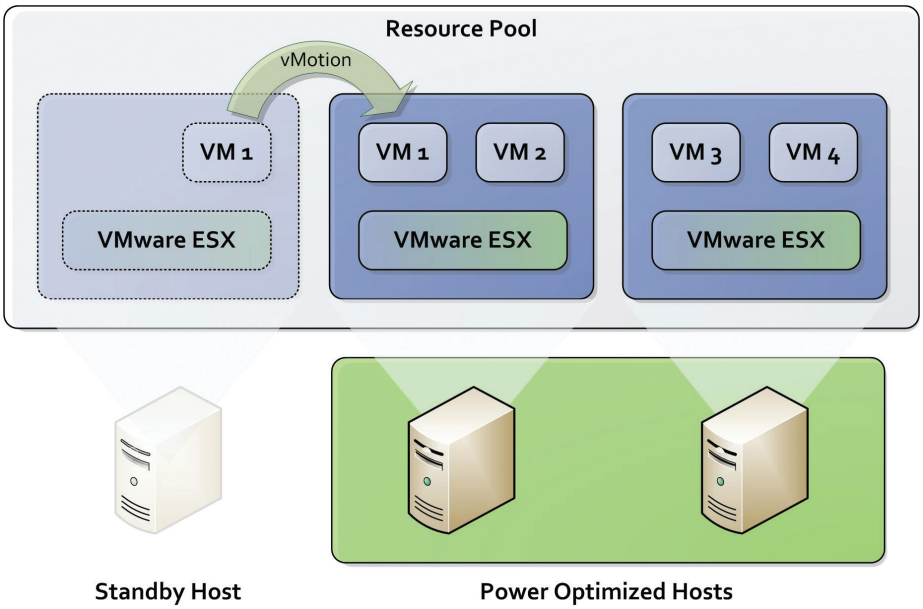

Figure 4: VMware Distributed Resource Scheduling

Figure 5 shows the DPM settings screen on VMware vCenter. As shown in the in the figure, it is possible to choose between three different dynamic management configurations:

- **Disabled (Off):** No power management is carried out by VMware vCenter. Even if some virtual machines have individual power management settings, these are not applied.
- **Manual:** VMware vCenter provides virtual machine migration and node power on/off recommendations in order to optimize energy consumption, but these recommendations must be manually approved by the platform administrator prior to their application.
- **Automatic:** VMware vCenter automatically migrates virtual machines and power servers on/off in order to optimize the platform's energy efficiency. In this configuration, users can also choose the desired automation level. The more conservative the automation threshold is, the higher the energy saving has to be to trigger the reduction of cluster capacity.

Apart from choosing a global configuration for DPM, it is also possible to customize these settings for individual virtual machines, as shown in Figure 6. On this screen, it is possible to select whether the virtual machine will follow the default policy or rather is applied in inactive, manual, or automatic configuration. As mentioned before, these individual settings are not taken into account if the global DPM configuration is set to "off."

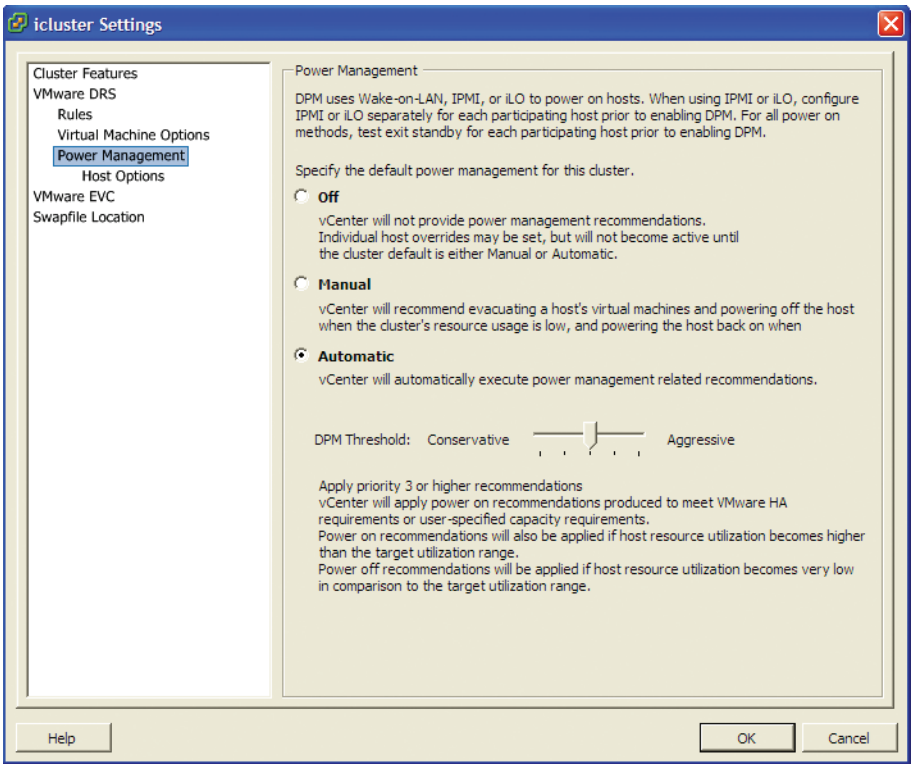

Figure 5: VMware Distributed Resource Scheduling

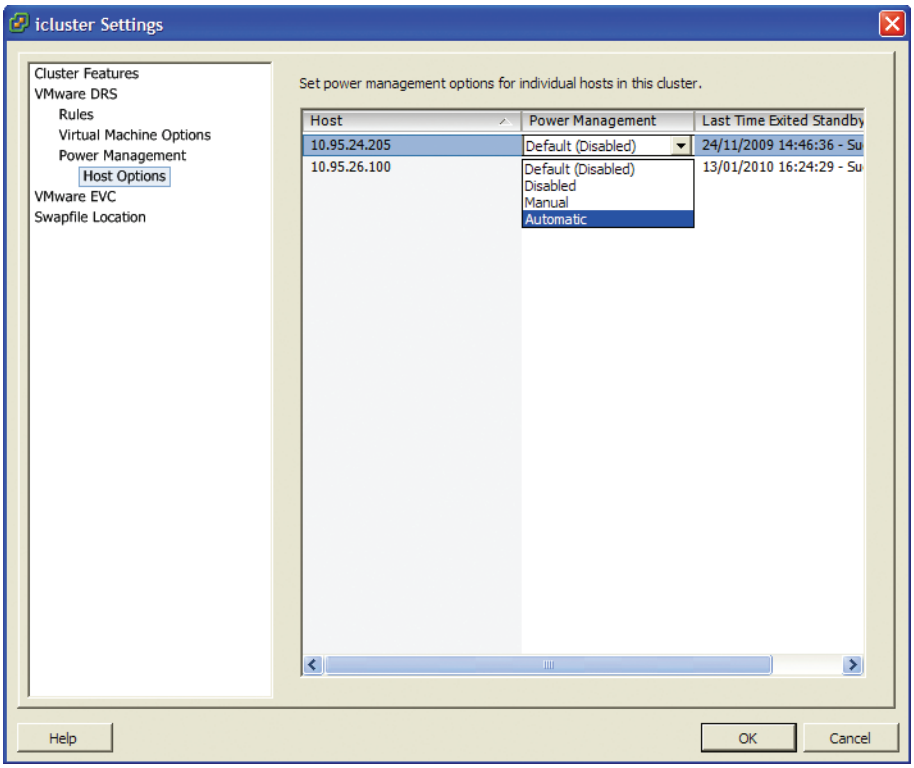

Figure 6: Setting Individual Host Options in VMware DPM

VMware vCenter* Interface

The VMware vCenter product includes a Java SDK that allows building third-party virtual infrastructure management applications. In order to operate the virtualized infrastructure and enable managing the APC nodes for application scale-in and scale-out, a group of Java libraries (VMware Wrapper libraries) have been developed in order to interact with vCenter. Specifically, the VMware Wrapper libraries implement the methods that allow executing the following actions:

- **Connect** with VMware vCenter.
- **Search** an entity managed by VMware vCenter.
- **Get an entity** object's (cluster, virtual machine, host, etc.) state and properties.
- **Get** and **update** the power state of a virtual machine (power on, power off, suspend, etc.).
- **Get** and **update** the power state of a host (power on, power off, enter maintenance state, etc.).
- **Migrate** a virtual machine (cold and live migration).
- **Get** and update a cluster's DRS and DPM properties.
- **Vertically** scale a virtual machine by managing its CPU and memory limits.
- **Reconfigure** a virtual machine's hardware (disk, network interfaces, CPU, etc.).

Apart from interacting with the virtualization layer through the VMware vSphere API exported by vCenter, it has also been necessary to interact directly with the ESX hosts supporting the virtualized infrastructure in order to obtain reliable CPU utilization measurements for dynamic infrastructure management. This necessity is motivated by the fact that, in the vCenter version used in our proof of concept, calls to the vSphere API return CPU usage information, which does not take into account the effect of hardware power management and hyper-threading technologies. As a workaround until future versions of vSphere allow to directly fetch such information from the vSphere API, it has been necessary to interact with ESX hosts to obtain CPU utilization metrics.

Additionally, a Java-based logging tool has been developed to obtain and log CPU utilization information from a series of ESX hosts. The tool has been used to obtain the CPU utilization records used in the use case evaluation activities.

Application Level Power Management

Application-level power management comprises the monitoring of service-specific KPIs, as well as service-level management actions such as controlling application scalability and workload placement.

In this proof of concept, application-level management consisted of KPI monitoring only, since application scalability and placement was delegated to the virtualization tier.

APC Monitoring Agents Interface

Service-specific KPI monitors and control capabilities in this proof of concept were based on the features supplied by Telefónica I+D's service monitoring agents.

In order to recover APC KPI statistics, an SSH Wrapper that allows connecting to the APC monitoring agents and remotely executing commands on them was implemented. Interfacing between the

Dynamic Infrastructure Manager and the APC monitoring agents was implemented by means of the wrapper library.

Target Scenarios for the Proof of Concept

The main goal of this initiative is to demonstrate the application of intelligent management of infrastructure assets for attaining significant energy savings and a more rational utilization of the available energy resources, while preserving the service level of the managed platforms or the user perceived quality of the services hosted on them.

Each scenario, namely enforce priority power allocation and energy management, was tested separately and then in combination. This Proof of Concept (PoC) illustrates the use of multiple concurrent policies, each carrying a different operational objective, namely enforcing priority allocation of limited power and a reduction in energy consumption. Two separate use cases were applied concurrently to cover different operational scenarios for the infrastructure under control:

- Power-aware dynamic reconfiguration for adjusting the infrastructure profile in real time to meet workload demand. The goal is in this case to fulfill the committed service level agreements (SLAs) with the minimum possible amount of resources. Power-aware targets facilities facing no power or CO2 emission limitations. In this case, the goal is optimizing energy efficiency, fulfilling the committed Service-Level Agreements (SLAs) with the minimum possible amount of resources.
- Power-constrained operation, the main goal of the dynamic management solution, is to enforce the overall power or CO2 emission restrictions of a facility. The emission restrictions translate into power-limited operation. Under these conditions, the dynamic manager apportions power to the different services. The apportionment policies enforced maximize SLA compliance while taking into account the different priorities of the services hosted on the managed site.

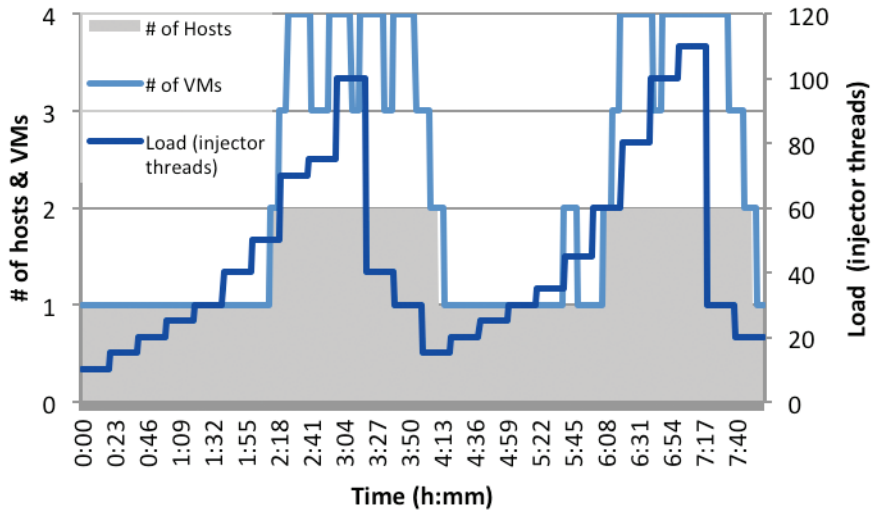

Figure 7: Power-aware Dynamic Reconfiguration Actions

Proof of Concept Experimental Results

In this section we present the results from experiments performed on the system described in the previous section arranged in two sub-sections: Power-aware dynamic reconfiguration and under power constrained operations.

Evaluation of Power-Aware Dynamic Reconfiguration

Assuming an environment where we operate under no power or CO2 emission limitations, this set of tests intends to evaluate the energy-saving benefits derived from the adoption of a dynamic reconfiguration policy to adaptively reconfigure the equipment to track large variations in workload demand and the impact on service performance of such management approach.

Dynamic reconfiguration alone does reduce the number of servers necessary to meet periods of low demand and improves OpEx performance by extending power proportional computing range of a pool of servers for the pool as a group. The pool of active servers is maintained at its most efficient operating range even during periods of low workload demand. Equipment not needed to meet workload demand is turned off. This operating procedure leads to significant energy savings over more traditional methods that rely on the autonomic power management capabilities in the CPU only.

Dynamic Reconfiguration

Figure 7 shows the first group of tests tackled in the dynamic reconfiguration scenario, where server parking to the S5 state and service architecture control were applied to the Gold server group according to energy-saving objectives. The figure shows how the dynamic energy manager controls the number of active servers and application nodes (i.e. VMs) depending on service demand in

order to maintain the service SLA within the desired range. In this trial experience, the KPI target was set to maintaining the daily average of the service response time under 300 milliseconds. The signaling load injected followed the pattern found on the production service instance in a typical working day, accelerated according to a 3:1 time scale to enable the execution of a daily cycle in eight hours.

Figure 8 depicts the effect of these management actions on service performance, as measured by service response time. As shown, the average daily response time increases from 264 milliseconds in the power unmanaged case to 299 milliseconds with dynamic reconfiguration. The dynamic reconfiguration results are still within the target SLA. Hence, the change is not really 14 percent degradation in service performance, but an adjustment up to the service level commitment.

Figure 9 allows quantifying the benefits introduced through the adoption of power-aware dynamic reconfiguration technologies and the controlled degradation in service performance discussed in Figure 8. As shown, the service power draw is now more proportional to the signaling load. The energy consumption measured during the execution of the test diminishes from 3694 watt-hours to 2683 watt-hours, which represents a 27 percent reduction.

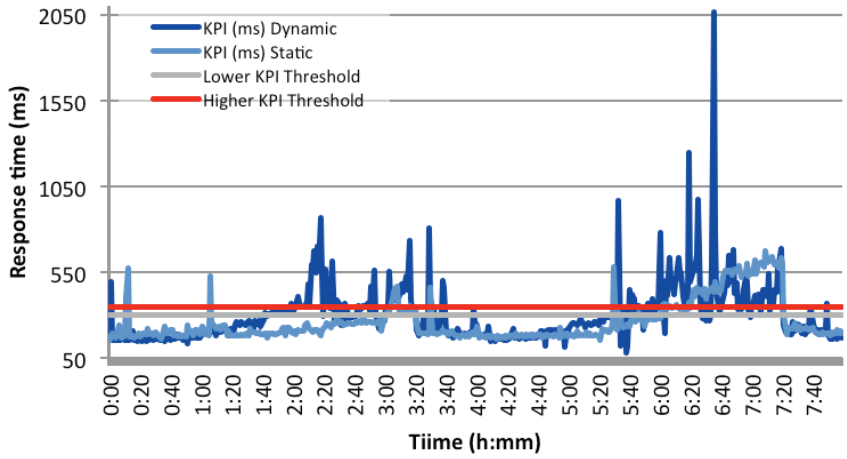

Figure 8: Service Response Time Comparison (Static vs. Dynamic)

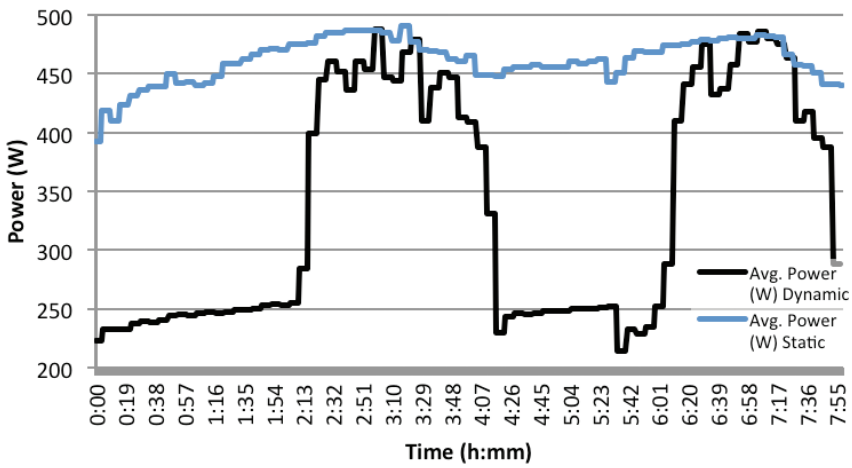

Figure 9: Server Group Power Draw Comparison (Static vs. Dynamic)

Evaluation of Power-Constrained Operation

A second group of tests addressed the power-constrained operation scenario. Server power capping was applied in order to enforce a facility power limitation of 800 watts, taking into account the service-aware considerations described in Section 2. As shown in Figure 10, the average response time of the high-priority service instance improves by 46 percent when compared to the static application of the power cap to the whole facility, thanks to the introduction of service-aware power shifting techniques. The tradeoff is that, by taking power out of the low-priority instance to apportion it to the high-priority one, the response time of the former increases by 52 percent.

If we consider the power consumption of the whole facility and the Gold and Silver server groups, we may better appreciate the effect of the service-aware power apportionment process. As shown in Figure 12, a total of 3190 watt-hours were consumed under static capping: 48 percent by the Gold server group and 52 percent by the Silver group. With a service-aware power management policy in effect, the relationship flips. In this case, a total of 3195 watt-hours were consumed, of which 53 percent was apportioned to the Gold server group and 47 percent to the Silver one.

Figure 11 shows the effect of the dynamic power capping and reconfiguration actions on the platform's power consumption. The overall facility power cap is enforced, applying also more stringent caps on low-utilization intervals to reduce the power consumption further. Dynamic caps are also applied on the server groups as mandated by service performance, making sure that the high-priority service instance is favored whenever a competition for energy resources takes place. These caps introduce the correlation between signaling load and energy consumption that may be observed in the figure.

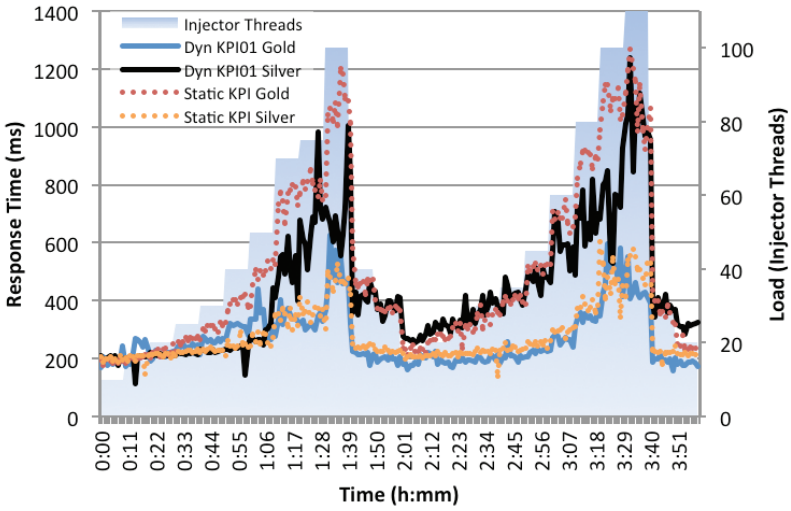

Figure 10: Service Performance (Dynamic vs. Static Scenario)

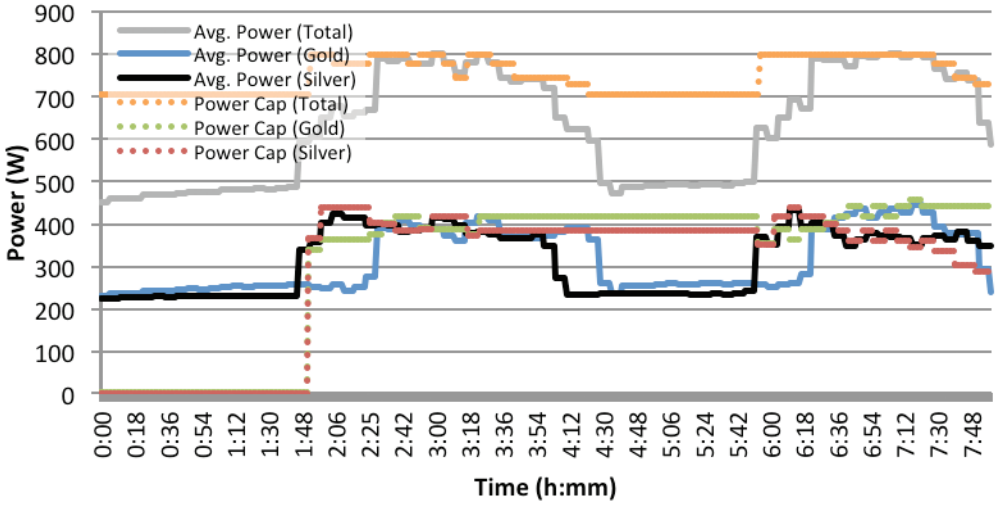

Figure 11: Power Consumption (Combined Scenario)

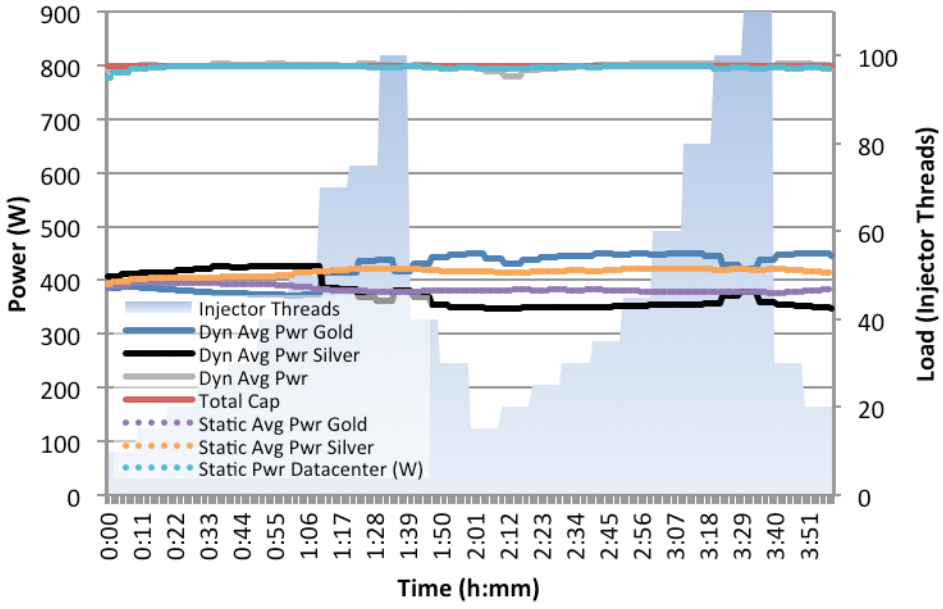

Figure 12: Power Consumption (Dynamic vs. Static Scenario)

Conclusion

As we mentioned in the introduction to this document, the main goal of this technology evaluation initiative was to demonstrate how the intelligent management of infrastructure assets may yield significant energy savings and a more rational utilization of the available energy resources with minimal impact on the service level of the managed platforms or the user-perceived quality of the services hosted on them.

According to the results obtained in our use case evaluation tests, we can consider the point well proven, since it becomes possible to achieve a notable reduction in energy consumption with a minimal and controlled reduction in service performance. In our tests, we were able to reduce the energy consumption by 27.35 percent with only a 13.56 percent increase in the service response time, ensuring that the average service response time objectives were met. When analyzing these results, it is worth taking two considerations into account. First, that energy savings are strongly dependent on the horizontal scalability features of the target service and the differences in load intensity and duration between the peak and valley load periods. Shorter and higher-intensity peak load periods in comparison to the average service load will result in bigger energy savings. The second consideration is that the dynamic infrastructure management policy allows selecting the desired trade-off between service performance and energy savings. The higher the allowed service performance degradation is, the bigger the energy consumption reduction will be.

In our power-constrained operation tests, we have been able to demonstrate how a data center-wide power cap is not a fair method for power budget apportioning. With an overall restriction, all services are treated equally, regardless of their priority level and the service level committed. It may even be the case that high-priority services are penalized in favor of standard services if these are more power hungry.

Our tests show how, by taking into account service priority and SLAs when applying power caps, it becomes possible to prioritize services as desired, favoring them up to the maximum desired degradation in standard service performance.

Our tests have also allowed us to observe that, when using Intel DCM to apply power caps to facilities and groups of servers, there are significant differences in behavior depending on how these caps are implemented. A detailed policy, where the overall cap is also fully apportioned to the different server groups by means of sub-policies, allows a tighter control over the power apportionment process. However, the lack of leeway on the facility-level policy may lead to violations of the global power cap whenever a lower-level policy cannot be maintained. By setting limitations in the facility as a whole and the low-priority assets only, we can ensure that facility power constraints will be strictly enforced. Still, the degree of control over the available power budget is lower than in the previous case.

The application of one mechanism or the other will mostly depend on whether the facility's power limitation is a hard or soft constraint. Another possible approach would be adopting a reactive policy, where alarm information is processed whenever a lower-level policy cannot be met. In those circumstances, the dynamic manager could shift the power budget to absorb the excess in power consumption of the non-abiding server group.

APPENDIX A: Server Power Management

Intel® Power Management Technologies

Microprocessors are possibly the most energy intensive components in servers and have traditionally been the focus of power management strategies. Emergent technologies such as solid state drives have the potential to significantly reduce power consumption and, in the future, management of memory power consumption may be incorporated.

Intel Intelligent Power Node Manager and Intel DCM are designed to address typical data center power requirements such as described above.

Intel Intelligent Power Node Manager is implemented on Intel® server chipsets starting with Intel Xeon processor 5500 series based platforms. Intel Intelligent Power Node Manager provides power and thermal monitoring and policy based power management for an individual server and is exposed through a standards based IPMI interface on supported Baseboard Management Controllers (BMCs). Intel Intelligent Power Node Manager requires an instrumented power supply that conforms to the PMBus standard.

Intel DCM SDK provides power and thermal monitoring and management for servers, racks, and groups of servers in data centers. Management Console Vendors (ISVs) and System Integrators (SIs) can integrate Intel DCM into their console or command-line applications to provide high value power management features. These technologies enable new power management paradigms and minimize workload performance impact.

Intel® Data Center Manager

From a data center perspective, the ability to regulate power consumption of just a single server has a small impact and is not intrinsically useful. Intel DCM provides the means to control servers as a group, and to monitor the power for the group of servers to allow meeting a global power target for that group of servers. This function is provided by the Intel DCM software development kit and shown in Figure 13.

Note that Intel DCM implements a feedback control mechanism very similar to the mechanism that regulates power consumption for a single server, but at a much larger scale. Instead of watching one or two power supplies, Intel DCM oversees the power consumption of multiple servers or "nodes" whose numbers can range up to thousands.

Figure 14, depicts an expanded view of the Intel DCM control loop as well as the relationship with the Intel Intelligent Power Node Manager control loop underneath. No specific agents need to run in each node. Intel DCM communicates with the board management controller (BMC) in each node for setting power targets and for doing readouts of the actual power consumed. Intel Intelligent Power Node Manager firmware takes care of ensuring that the individual server meets the assigned power consumption target.

Intel DCM was purposely architected as an SDK as a building block for industry players to build more sophisticated and valuable capabilities for the benefit of data center operators. Some Intel Intelligent Power Node Manager-enabled servers come with inlet temperature sensors. This allows the application to monitor the inlet temperature of a group of servers, and if the temperature rises above a certain threshold, it can take a number of measures, from throttling back power consumption to reducing thermal stresses.

An application interfacing with Intel DCM no longer "sees" individual server nodes; the application code's power policy engine designates a power consumption target to Intel DCM through the Intel DCM API. Intel DCM in turn breaks down the power target into power targets to the individual nodes under its command.

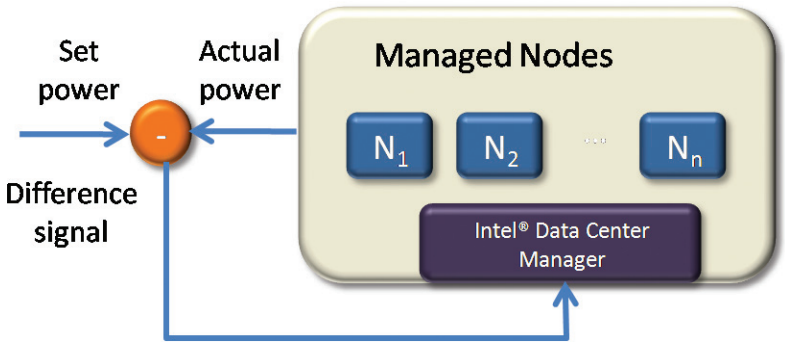

Figure 13: Power control loop in a group managed by Intel DCM

Intel® Intelligent Power Node Manager

Intel Xeon processors regulate power consumption through voltage and clock frequency scaling. Reduction of the clock frequency reduces power consumption, as does lowering voltage. The scale of reduction is accomplished through a series of discrete steps, each with a specific voltage and frequency. The Intel Xeon processor 5500 series can support 13 power steps. These steps are defined under the ACPI08 standard and are colloquially called P-states. P0 is nominally the normal operating state with no power constraints. P1, P2, and so on aggressively increase the power capped states.

Voltage and frequency scaling also impacts overall system performance, and therefore will constrain applications. The control range is limited to a few tens of watts per individual microprocessor. This may seem insignificant at the individual microprocessor level, however, when applied to thousands or tens of thousands of microprocessors typically found in a large data center, potential power savings amount to hundreds of kilowatt hours per month. Intel Intelligent Power Node Manager is a chipset extension to the BMC that supports in-band/out-of-band power monitoring and management at the node (server) level. Some of the key features include:

- Real-time power monitoring
- Platform (server) power capping
- Power threshold alerts

Figure15 below shows the Intel Intelligent Power Node Manager server power management closed control loop.

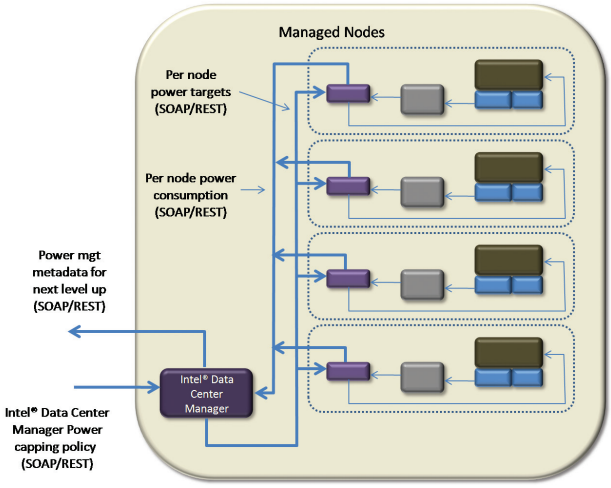

Figure 14: Intel DCM power control loop, detailed view

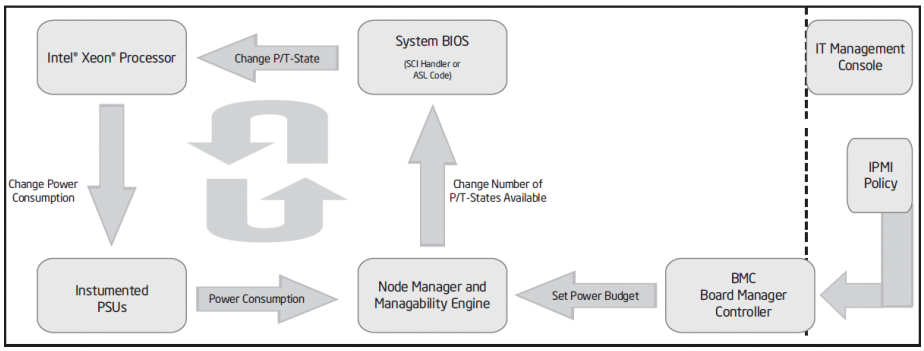

Figure 15: Intel Intelligent Power Node Manager server power management closed control loop.

Disclaimers

Δ Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. See www.intel.com/products/processor_number for details.