

## TOWARDS ENERGY EFFICIENCY IN COMPUTING

# HACIA LA EFICIENCIA ENERGÉTICA EN LA COMPUTACIÓN

**Bernardo Aebischeri**

*Instituto Suizo de tecnología. Zurich, Switzerland  
bernabis@europe.com*

(Tipo de Artículo: **REFLEXIÓN**. Recibido el 04/09/2010. Aceptado el 30/10/2010)

**Abstract** – *Today, almost everyone is aware of the energy problem facing humanity: our primary energy sources are being depleted, while demand, both commercial and domestic, is increasing. Moreover, the side effects of using traditional energy sources bring significant global environmental considerations. This paper offers an answer to the question: What is the role of computers in energy demand, and where we need to focus to reduce consumption and improve energy efficiency?*

**Keywords:** *Energy efficiency, computing, power, energy.*

**Resumen** – Actualmente, casi todo el mundo es consciente del problema energético que enfrenta la humanidad: nuestras fuentes primarias de energía se están agotando, mientras que la demanda, tanto en entornos comerciales como en los domésticos, es cada vez mayor. Por otra parte, los efectos secundarios del uso de las fuentes tradicionales de energía traen importantes consideraciones ambientales globales. En este trabajo se ofrece una respuesta a la pregunta: ¿Qué papel juegan los equipos informáticos en la demanda de energía, y dónde debemos focalizarnos para reducir el consumo y mejorar la eficiencia energética?

**Palabras clave:** Eficiencia energética, computación, potencia, energía.

### 1. INTRODUCCIÓN

La expectativa por desarrollar nuevas fuentes de energía sostenible tiene por lo menos tres décadas de trabajo constante. Steve Chu [1], secretario de Energía de los EE.UU., puso esta situación en contexto: "Se necesita una estrategia doble para resolver el problema de la energía: 1) maximizar la eficiencia energética y disminuir el consumo de energía; y 2) desarrollar nuevas fuentes de energía limpia. El primero seguirá siendo el fruto más bajo en el árbol para las próximas décadas".

¿Qué papel juegan los equipos informáticos en la demanda de energía, y dónde debemos focalizarnos para reducir el consumo y mejorar la eficiencia energética?

En agosto de 2007, la Agencia de Protección Ambiental EPA [2] le envió al Congreso de los EE.UU. un reporte acerca de la eficiencia energética de los servidores y centros de datos. Algunas de las conclusiones clave del informe son:

- Los servidores y centros de datos consumieron 61 billones de kWh en 2006, lo que representó el 1.5% del consumo total de energía ese año en los EE.UU., y US\$ 4.5 billones en costos; en promedio equivalente a las necesidades energéticas de 5.8 millones de hogares.
- La electricidad consumida en este sector se duplicó entre 2000 y 2006, una tendencia que se espera continúe.
- Los sistemas de infraestructura necesarios para apoyar la operación de equipos IT –por ejemplo equipos de protección y sistemas de refrigeración– también consumen una cantidad significativa de potencia, que abarca el 50% del consumo anual de energía.

Extractos de este informe se muestran en la Figura 1 y en la Tabla 1, y en ellos sobresalen especialmente dos puntos: el primero se refiere a que la misma infraestructura de los espacios para funcionamiento del equipo informático consume mucha energía. Este consumo de la infraestructura está representado principalmente en la calefacción, la ventilación y el aire acondicionado; además, la que se utiliza para convertir y transmitir potencia y mantener su continuidad –este último incluye los transformadores y los equipos de conmutación de alimentación y transmisión, así como el equipo de acondicionamiento de potencia y las fuentes de alimentación ininterrumpida. Este factor es de gran importancia, pero no puede ser el área principal a la que se enfrenten los profesionales de la informática.

El segundo punto que sobresale en el informe es el equipo de cómputo en sí. Sólo el volumen de servidores, de los cinco tipos de equipos IT estudiados, fue responsable de la mayor parte de energía utilizada: 68%. Suponiendo que la tasa de crecimiento anual –*Compound Annual Growth Rate CAGR*– del volumen de servidores se mantiene en 17%, se convertirán en el principal objetivo para reducir energía en el área de IT. En el mismo informe se detalla que la tasa de crecimiento de los dispositivos de almacenamiento es de 20% –una tasa de la que los datos más recientes sugieren que se está acelerando–, lo que indica otra tendencia importante.



Fig. 1. Consumo de electricidad de los componentes de uso final 2000-2006

Fuente: Environmental Protection Agency

**TABLA 1**  
Electricidad que consumen los componentes de uso final 2000 a 2006

Componente	2000		2006		2000/2006
	Uso de energía	%	Uso de Energía	%	Uso de energía
Infraestructura	14.1	50	30.7	50	14
Equipo de redes	1.4	5	3.0	5	14
Almacenamiento	1.1	4	3.2	5	20
Servidores grandes	1.1	4	1.5	2	5
Servidores medios	2.5	9	2.2	4	-2
Volumen servidores	8.0	29	20.9	34	17
<b>Total</b>	<b>28.2</b>		<b>61.4</b>		<b>14</b>

Fuente: Environmental Protection Agency

Si el crecimiento exponencial de los equipos informáticos de los centros de datos, revelado en este estudio, se mantiene, se espera que para el 2011 se necesite aproximadamente el doble de la energía que se demandó en el 2006 para dichos centros. Esto plantea desafíos obvios más allá de los económicos. Por ejemplo, la demanda máxima instantánea se espera que aumente de 7GW en 2006 a 12GW en 2011, por lo que se necesitarán diez nuevas plantas de potencia base para satisfacer esta demanda.

En algunas áreas, las limitaciones físicas en disponibilidad de potencia son ya una limitante para los centros de datos; un gerente gestor de TI en Morgan Stanley [3] observó no hace mucho que la empresa ya no era capaz físicamente de obtener la potencia necesaria para hacer funcionar un nuevo centro de datos en Manhattan. La situación es tan grave, que corporaciones como eBay, Google, Amazon, Microsoft y Yahoo están buscando lugares adecuados en los que puedan construir centros de datos, necesarios para correr sus aplicaciones web y ofrecer los servicios que actualmente poseen.

Algunas de estas empresas ya han negociado con ciertos estados en los EE.UU., así como a nivel internacional, para la construcción de estas instalaciones además de las plantas de poder necesarias para su suministro. Hace unos años, Google provocó lo que algunos periodistas denominaron "una moderna carrera armamentista" cuando situó una nueva instalación a lo largo del río Columbia en Washington. Los beneficios combinados de bajos costos en tierras, bajas

temperatura en el ambiente exterior, y la disponibilidad de agua potable para la refrigeración y la generación de potencia hidroeléctrica, podría proporcionarle a Google un antídoto para los agudos problemas de disponibilidad de energía y los costos asociados.

Existe alguna evidencia<sup>1</sup> de que la cantidad de energía consumida por los equipos informáticos móviles y de escritorio es aproximadamente de la misma magnitud que la utilizada por los servidores en los centros de datos, aunque no se tiene un estudio completo y autorizado al cual referirse. Los datos de la EPA que aquí se presenta ofrecen una perspectiva detallada sobre dónde va la energía, en el panorama de la computación, en el cada vez más importante segmento de los servidores. Además, alguna fundamentación tendrá esta información, cuando el programa EnergyStar® de la EPA en electrónica de consumo puso, como resultado al anterior informe, a la computación móvil y a la de escritorio en el mismo espacio que el resto de componentes informáticos.

## 2. ADMINISTRACIÓN DE POTENCIA EN LA INFORMÁTICA ACTUAL

Quizás el factor clave a considerar con los sistemas informáticos actuales es que la cantidad de potencia que consumen no se ajusta suficientemente a la cantidad de trabajo que el sistema realiza. El objetivo principal de diseño para la mayoría de los sistemas informáticos de

<sup>1</sup> La Administración de Información Energética de los EE.UU. informó, en 2001, que los computadores e impresoras en los hogares americanos consumieron 23.1 TW/h [6]; cifra similar en 2006 [7].

propósito general, hasta la fecha, ha sido el de maximizar el rendimiento –o, tal vez, el rendimiento a un precio determinado– con muy poca consideración en cuanto al uso de la energía. Esto ha cambiado rápidamente a medida que nos acercamos al punto en el que el costo de un equipo de cómputo será superado por el costo de la energía para su operación, incluso durante su relativamente corto –3 a 5 años– período de amortización; es el momento de prestar la suficiente atención a la energía en el diseño del sistema.

Aunque el ejercicio se ha hecho para calcular la energía proporcional [4] –es decir, la cantidad de potencia necesaria se corresponde directamente con el grado en que el sistema o componente se utiliza–, está muy lejos de la situación actual. Muchos componentes de los sistemas informáticos actuales tienen muy bajos niveles de utilización, y la mayoría de los sistemas gastan gran parte de su tiempo de operación en niveles de uso relativamente bajos.

Las fuentes de potencia han sido notorias por su ineficiencia, especialmente con una carga baja, y los ventiladores pueden desperdiciar mucha energía cuando se utilizan sin cuidado. Sin embargo, en los últimos cuatro años la eficiencia de las fuentes de alimentación ha mejorado [5]. De hecho, han surgido algoritmos que ajustan más continuamente las velocidades del ventilador en relación con las necesidades térmicas, en lugar de utilizar sólo algunos puntos de velocidad discreta. Sin embargo, la mayoría de los componentes de hardware en los sistemas informáticos de hoy todavía se deben gestionar de forma explícita, y las concepciones y facilidades para administrar la potencia en los sistemas informáticos, ampliamente desplegadas, aun son rudimentarias.

## 2.1 Administración de la potencia

Existen dos modalidades básicas de administración de potencia: 1) funcionamiento vs. suspensión, en la que un componente o sistema se puede apagar cuando no esté en uso –es decir, una vez que cambie su estado a en reposo–, pero se vuelve a encender cuando sea necesario; y 2) de ajuste de rendimiento –mientras está en ejecución–, en la que el nivel de rendimiento de un componente puede bajar o subir con base en el nivel de utilización observado u otra necesidad de la carga de trabajo.

La opción de funcionamiento vs. suspensión a menudo se llama “estados de energía”, de un componente o sistema. Si bien existe un estado único para representar el funcionamiento, puede haber más de uno para la suspensión, lo que permite que la potencia se suspenda más progresivamente, si existe alguna estructura de potencia relevante para su implementación, desde el hardware asociado al componente o sistema.

Las CPUs, por ejemplo, pueden suspender su ejecución simplemente al detener la emisión de instrucciones o al apagar su circuito de reloj. Sin embargo, algunos estados de potencia más profundos pueden remover sucesivamente la energía del caché de los procesadores, de los TLBs –Translation Lookaside Buffers–, de los controladores de memoria, y así sucesivamente. Mientras más componentes de hardware se suspendan más potencia se ahorra, pero entonces habrá una latencia mayor para reanudar su funcionamiento, o será necesaria energía adicional para salvar y restaurar el contenido del hardware al reiniciarlo o para ambos.

Las opciones de ajuste de rendimiento, mientras se ejecutan, se conocen más naturalmente como “estados de potencia” de un componente. Una técnica ampliamente aplicada para ajustar el rendimiento es la de cambiar la frecuencia de operación del componente. Cuando la velocidad del reloj es baja, los niveles de voltaje de operación también se pueden reducir, y estos dos factores juntos –normalmente llamados Dynamic Voltage and Frequency Scaling DVFS– se traducen en un ahorro de potencia compuesto. Los estados de rendimiento se introdujeron por primera vez para las CPUs, ya que los procesadores se encuentran entre los consumidores más importantes de potencia en la plataforma hardware –un CPU multi-núcleo moderno está en el rango de 35W y 165W. Los estados de rendimiento también se pueden utilizar para controlar el tamaño de caché activo, el número y/o tasas de operación de la memoria, las interconexiones I/O, y otras similares.

La arquitectura más ampliamente implementada para la administración de potencia es la Advanced Configuration and Power Interface ACPI, que ha evolucionado junto con la arquitectura Intel, basadas en las plataformas hardware más ampliamente disponibles para CPUs y sus componentes relacionados. Aunque hay muchos aspectos detallados para la especificación, ACPI ofrece principalmente los controles necesarios para implementar las dos modalidades de administración de potencia que se acaban de describir. Además, define los estados de potencia: siete a nivel de todo el sistema, llamados estados-S –S0 a S6–, y cuatro a nivel de cada dispositivo llamados estados-D –D0 a D3. El estado 0 –S0 para el sistema y D0 para cada dispositivo– indica estado en ejecución o activo, mientras que los de mayor número son los estados no ejecutados o inactivos con potencia sucesivamente más baja –y en consecuencia niveles decrecientes de disponibilidad o de disposición de ejecución. ACPI también define estados de rendimiento, llamados estados-P –P0 a P15, permitiendo un máximo de 16 por dispositivo–, que afectan el rendimiento operativo del componente durante la ejecución. Ambos afectan el consumo de potencia.

### 3. EFICIENCIA ENERGÉTICA EN LA COMPUTACIÓN

A pesar de que ACPI es un importante estándar con amplio apoyo de parte de los fabricantes, sólo proporciona un mecanismo para controlar los aspectos del sistema con el propósito de afectar su consumo de potencia. Esto permite pero no establece explícitamente la eficiencia energética. Se necesitan aspectos de más alto nivel de la arquitectura general del sistema para explotar este o cualquier otro mecanismo similar.

¿Qué tanto difiere el cálculo de la eficiencia energética de la administración de potencia? ¿Cómo se puede saber que se ha resuelto el problema de la eficiencia energética para un sistema informático? He aquí una visión simple: "El sistema consume la cantidad mínima de energía<sup>2</sup> requerida para realizar cualquier tarea". En otras palabras, la eficiencia energética es un problema de optimización. Dicho sistema deberá ajustar dinámicamente sus recursos hardware, de forma que sólo se habiliten los que se necesita para realizar esas tareas –ya sea para completarlas a tiempo, o por analogía, para proporcionar el rendimiento requerido para mantener un nivel de servicio establecido–, y como un resultado, la energía total utilizada se reducirá al mínimo.

Tradicionalmente, los sistemas se han diseñado para obtener el máximo rendimiento para la carga de trabajo. En los sistemas energéticamente eficientes, en algunos casos todavía se desea obtener el máximo de rendimiento para algunas tareas o para la carga de trabajo completa, pero el sistema también debe reducir al mínimo el uso de energía. Es importante entender que el rendimiento y la eficiencia energética no son mutuamente excluyentes. Por ejemplo, incluso cuando se logra el máximo rendimiento, un recurso que se deba desactivar o cuyo rendimiento individual se pueda reducir sin afectar el tiempo para completar lo mejor posible la carga de trabajo o el rendimiento, constituyen optimización energética.

De hecho, en cualquier sistema existen pocas –o ninguna– situaciones en las que se explote la capacidad total de los recursos hardware –es decir, todos operan a sus niveles de máximo rendimiento. Los sistemas que se esfuerzan por obtener el máximo rendimiento en todo momento son notoriamente sobre-saturados –y en consecuencia poco utilizados. Sin embargo, las personas que participan en el diseño de sistemas informáticos prácticos pueden notar que nuestra ciencia es débil en esta área –área que se podría

llamar "planificación de la capacidad dinámica y del aprovisionamiento dinámico".

La optimización de la energía obviamente está sujeta a ciertas restricciones. Algunos ejemplos son:

#### 3.1 Se deben mantener los niveles de rendimiento requeridos

*Las tareas con plazos se deben completar a tiempo.* En el caso general, se especifica una fecha límite para una tarea o la carga de trabajo. Cuando se especifica una fecha límite que es menor o igual a la óptima en la que el sistema puede lograr ese trabajo con cualquiera o con todos los recursos hardware, implica también un máximo rendimiento. Este es efectivamente un caso extremo.

El máximo rendimiento para una tarea o carga de trabajo proporciona una estipulación implícita de la fecha límite óptima, o "tan pronto como sea posible"<sup>3</sup>. En este caso, la optimización de energía se limita a los recursos que se pueden desactivar o cuyo rendimiento individual se puede reducir, sin afectar el tiempo para completar lo mejor posible la carga de trabajo.

Si se especifica una fecha límite menor que la mejor fecha límite alcanzable, el cálculo puede tardar algún tiempo hasta esta fecha, y el sistema puede buscar un mínimo de energía más global para la tarea o trabajo. Las fechas límite podrían considerarse "difíciles", en cuyo caso el asignador de recursos para optimizar energía del sistema debe garantizar de alguna manera su cumplimiento –lo que plantea problemas de difícil aplicación–; o considerarse "suaves", en cuyo caso sólo se puede tolerar un mayor esfuerzo.

*Los servicios deben operar de acuerdo con el volumen de trabajo.* Para los servicios en línea, la noción de volumen de trabajo, con el fin de caracterizar el nivel de rendimiento requerido, puede ser más adecuada que lograr un plazo de ejecución. Dado que los servicios, en su implementación, en última instancia pueden descomponerse en tareas individuales a completar, se espera que haya una técnica análoga –aunque los medios más adecuados de especificar su restricción de rendimiento puedan ser diferentes.

#### 3.2 El sistema debe responder a los cambios de la demanda

*Las cargas de trabajo reales no son estáticas.* La cantidad de trabajo establecido y los recursos

<sup>2</sup> Potencia es la relación de transferencia de energía respecto al tiempo, de modo que para una potencia constante, energía = potencia x tiempo. Potencia y energía son conceptos diferentes y no deben confundirse.

<sup>3</sup> Todos los valores de vencimiento  $D = t_i$ , menores que el más corto plazo a lograr, son equivalentes a establecer:  $D = t_0$ ; es decir:  $\{\forall t_i < t_0, [D = t_i] \approx [D = t_0]\}$ . Por lo tanto, se puede denotar el máximo rendimiento para  $D = 0$ .

necesarios para alcanzar un determinado nivel de rendimiento variarán mientras se ejecutan. La respuesta dinámica es una consideración práctica importante relacionada con el nivel de servicio.

*El volumen de trabajo (T) debe ser alcanzable dentro de la latencia (L).* La especificación de la latencia máxima en la que se puede activar la capacidad del hardware reservado o incrementar su nivel de rendimiento parece un requerimiento claro, pero también debe estar relacionada con las necesidades de rendimiento de la tarea o del trabajo en cuestión.

*El volumen de trabajo depende del tipo de tarea.* Una métrica como TPS –Transactions Per Second– puede ser relevante para la operación de sistemas de bases de datos, cuadros por segundo para el componente de renderización de un subsistema de generación de imágenes, o medidas correspondientes para un servicio de clasificación, interconexión I/O, o interfaz de red. El uso interactivo impone criterios de respuesta como medios de distribución en tiempo real: computacional, de almacenamiento, y de capacidad I/O requeridos para cumplir las tasas de ejecución de audio y video. Aquí se sugiere una manera por medio de la cual podría, en la práctica, manejarse tan diversos requisitos de volumen de trabajo.

*La potencia instantánea nunca debe exceder su límite (P).* Se puede especificar un límite de potencia máxima para respetar sus límites de disponibilidad prácticos –ya sea para un sistema individual o para un centro de datos en conjunto. En algunos casos, puede ser permisible exceder brevemente este límite.

Las combinaciones de estas restricciones significa que se debe esperar el exceso de restricción en algunas circunstancias, y por lo tanto también se requerirá una política para restringir la relajación. Podría ser preferible una precedencia estricta de restricciones o hacer entre ellas un intercambio más complejo.

#### 4. UNA PROPUESTA DE SOLUCIÓN

Teniendo en cuenta este concepto para calcular la eficiencia energética, ¿cómo deberían construirse los sistemas? ¿Cómo se esperaría que operara un sistema de energía eficiente?

Un sistema debería tener tres aspectos principales que podrían resolver este problema:

1. Debe ser capaz de construir un modelo de poder que le permita saber cómo y dónde se consume potencia, y cómo la puede manipular –este componente es la base para la promulgación de cualquier forma de administración de potencia.

2. Debe tener un medio para determinar los requisitos de desempeño de las tareas o de la carga de trabajo –ya sea por observación o por algún medio de comunicación más explícito. Estas son las limitaciones –de determinación y rendimiento– de la evaluación de componentes.
3. Por último, debe implementar un *optimizador de energía* –un medio para decidir una energía–, una configuración eficiente del hardware para todo momento que esté en funcionamiento. Una optimización que puede ser relativa –decidida heurísticamente– o absoluta –basada en técnicas analíticas. Esta es la capacidad –de planificación y dinámica– del componente de aprovisionamiento.

El primer aspecto es relativamente fácil de construir. El tercero es sin duda inmediatamente accesible, especialmente cuando la(s) técnica(s) de optimización se basa(n) en métodos heurísticos. La segunda consideración es la de mayor proporción, y representa una importante consecuencia perjudicial para el cálculo de la eficiencia de la energía, y podría demandar bases más formales –programación– para comunicar al sistema los requisitos de la carga de trabajo. Para esto se necesita una descripción de las necesidades básicas de aprovisionamiento de la carga de trabajo, junto con una forma de indicar tanto sus requisitos de rendimiento como el rendimiento en el momento.

También podría ser útil una forma de indicar a priori la sensibilidad esperada de los diferentes recursos del sistema a los cambios en el aprovisionamiento. Afortunadamente, existe una serie de enfoques prácticos para seguir la eficiencia energética antes de habilitar los refinamientos esperados.

#### 4.1 Modelo de potencia

Con el fin de administrar la eficiencia energética del hardware del sistema<sup>4</sup>, éste debe conocer los detalles específicos de potencia de los dispositivos físicos bajo su control. Los componentes de potencia gestionables deben desplegar el control que ofrecen en el modelo arquitectónico ACPI), tales como sus estados de potencia y de rendimiento, respectivamente –estados-D y estados-P–. Sin embargo, para permitir el modelado de potencia en relación con el rendimiento y la disponibilidad –es decir, en relación con su respuesta de activación–, la interfaz del componente también debe describir como mínimo lo siguiente:

<sup>4</sup> "El sistema", de forma natural, se refiere al sistema operativo, aunque está claro que también debe incluir al hipervisor para sistemas virtualizados. Se puede esperar que este concepto tendrá que ampliarse razonablemente para incluir algunos aspectos del *firmware* e incluso componentes de hardware y tiempos de ejecución importantes, como la máquina virtual Java, que es la responsable de la asignación de recursos.

- *El consumo de potencia* por estado para cada estado inactivo, o el rango de potencia para cada estado activo.
- *La latencia de transición de estados*, o el tiempo necesario para realizar cada transición de estado.
- *Energía para el cambio de estado*, o la energía consumida al cambiar de estado.

Una vez que el sistema tiene un modelo de potencia, consistente de todo el hardware que gestiona potencia, se tiene el fundamento básico para operacionalizar la optimización de energía. Es importante destacar que en él se tiene el conocimiento de los componentes que consumen más potencia y los que tienen los controles de respuesta de más alta capacidad que se puede utilizar para afectar el uso de potencia.

#### 4.2 Restricciones de carga de trabajo y de evaluación de desempeño

En su deseo de limitar la cantidad de hardware activo y reducir su rendimiento para minimizar el consumo de potencia, ¿cómo debe ser un sistema para saber si las tareas en ejecución todavía alcanzan el volumen de trabajo para mantener niveles adecuados de servicio o para lograr sus plazos?

La evaluación de la carga de trabajo está sujeta a la tarea o aplicación en cuestión. El sistema operativo puede observar el grado en que sus distintos recursos han sido y son utilizados, y podría utilizar estas observaciones como sus mejores bases para predecir las necesidades de recursos futuras, tanto para reducir como para ampliar las que tiene disponibles. Se trata de una base relativamente débil para determinar qué carga de trabajo se necesitará, especialmente para anticipar su sensibilidad de respuesta dinámica. Como resultado, el sistema tendrá que ser mucho más conservador en cuanto a su reducción de los recursos disponibles o sus niveles de rendimiento. Parece claro que el mejor resultado se logrará si las aplicaciones evalúan su carga de trabajo en relación con sus requisitos de nivel de servicio o plazos de ejecución, y que puedan transmitir esa información al sistema operativo a través de una interfaz. Entonces, el sistema puede utilizar esta información para hacer potencialmente ajustes de recursos mucho más agresivos y en consecuencia aplicar una solución de optimización de energía totalmente mejorada.

He aquí la dicotomía crucial: el sistema es responsable de resolver el problema de la optimización de energía de acuerdo con los recursos que asigna, mientras que la aplicación es responsable de monitorear su propio nivel de rendimiento e informar al sistema para que sean puestos a su disposición los recursos apropiados.

#### 4.3 Optimización de la energía por parte del Sistema

Una vez que conoce las características de potencia del hardware, y posiblemente con la información descriptiva acerca de sus limitaciones a partir del nivel de aplicación del software, el sistema operativo debe comenzar el proceso dinámico de ajustar el rendimiento del hardware y los niveles de disponibilidad para controlar el consumo de potencia y mejorar el uso de energía de todo el sistema. ¿Cómo puede el sistema operativo tomar tales decisiones?

*Métodos heurísticos.* Aprovisionar energía para una carga de trabajo máxima puede, en algunos casos, optimizarla. Esta es la conjetura según la cual “el rendimiento (máximo) es verde”, que se refleja en las ideas de carrera al ocio o carrera al sueño [8]. Aunque hay evidencia de que este enfoque tiene mérito del lado del cliente cuando el sistema está inactivo –especialmente para los sistemas embebidos y móviles donde se puede ahorrar hasta el 95% de la energía si todo el sistema se pone en un estado de suspensión–, no es claro cómo se puede aplicar para calcular del lado del servidor. En algunos casos existe un incremento no lineal de la potencia requerida para obtener aceleración lineal –carga de trabajo, por ejemplo el modo turbo de los procesadores contemporáneos de Intel– y por lo tanto, en todos los casos, la energía óptima no se encuentra en un punto de aprovisionamiento y rendimiento acorde con la máxima carga de trabajo.

Ajustar dinámicamente el nivel de rendimiento del hardware, en función de su utilización, es una heurística utilizada ampliamente para el mejoramiento energético en sistemas activos: hacia abajo con una utilización baja o hacia arriba con una utilización alta –utilización por debajo o por encima de cierto umbral para cierta duración. Esta puede ser una técnica eficaz, pero se limita a situaciones en las que, para hacer el cambio de estado, tanto la latencia como la energía son tan bajas que son insignificantes.

*Optimización basada en restricciones para un enfoque.* En algunos casos puede ser posible ofrecer una solución analítica completa para simplificar el problema a tal grado. Por ejemplo, si se considera solamente una tarea individual en una CPU individual con una compensación de poder/rendimiento bien comprendida, es relativamente sencillo especificar completamente una agenda en la que la sea posible unir la tarea a su fecha de entrega con un mínimo de energía total. De forma más general los resultados formales también son posibles [9]. Sin embargo, esto se basa en una serie de supuestos que a menudo no se tienen en la práctica, como buenas estimaciones del trabajo total requerido por un proceso. Supuestos más débiles requieren algoritmos de optimización en línea para lograr la

programación consiente de energía. Existen algunos trabajos en esta área, pero aún no los suficientes como para sostener un sistema operativo de propósito general [10].

Para que un enfoque basado en optimización sea aplicable de forma general, se requiere una serie de técnicas. En los casos más sencillos, es posible el funcionamiento autónomo a nivel de dispositivo; por ejemplo, una GPU –*Graphics Processing Unit*– a nivel de hardware, puede disminuir potencia sin usar instalaciones de hardware agresivamente, con base exclusivamente en las apreciaciones instantáneas de sus niveles de utilización, debido a que mientras sea necesaria la latencia para hacer una copia de seguridad de esas instalaciones, será intrascendente. Prácticas similares parecen ser aplicables en el uso de los estados-P de la CPU –rendimiento de la CPU y el ajuste costo-energía basado en la escala de voltaje y frecuencia–, ya que tanto la energía de transición de estados como la latencia son muy bajas.

Los cambios de estado de hardware que afectan la potencia pero presentan una latencia mucho mayor, y/o una cantidad de energía mucho mayor para hacer el cambio de estado, requieren un tratamiento diferente. Un ejemplo obvio es un disco duro desacelerándose, situación en la que hay que tener en cuenta la latencia extensa que se requiere para volverlo a iniciar; pero esta latencia no es la única preocupación. Los sistemas de memoria de los semiconductores en las que parte de la memoria física total podría apagarse si no es requerida, y donde la latencia de encendido puede ser cercana a cero, todavía tendrán una energía de transición resultante, ya que se puede requerir muchas transacciones en memoria para reunir el conjunto de trabajo en esas páginas físicas que permanecen activas<sup>5</sup>. Los recursos de esta clase requieren, para garantizar que la latencia de activación puede ser tolerada o mantenida y que la energía de cambio de estado será superada por la energía que se ahorrará mientras esté en ese estado, un mayor conocimiento de la tarea o el comportamiento de la carga de trabajo, así como un tratamiento anticipatorio de los recursos de hardware requeridos.

Algunas técnicas de optimización comunes se pueden basar en la latencia de cambio de estado, sus demandas de energía, y así sucesivamente, y de éstas podría surgir una taxonomía de las mismas –unas formales o analíticas, otras basadas en métodos más numéricos o heurísticos.

Aunque se espera que las técnicas específicas, para la apropiada optimización de energía para

<sup>5</sup> Es interesante considerar si las heurísticas tradicionales tales como la regla de cinco minutos, diseñada para optimizar por rendimiento la jerarquía de memoria, podría tener analogías en la optimización de energía.

recursos hardware diferentes o subsistemas, sea algo diferente y que estén sujetas a las propiedades de los recursos hardware en cuestión, la esperanza es que la composición de optimizadores de eficiencia energética para todos los recursos se acumularán para formar un esquema de eficiencia para el sistema en su conjunto<sup>6</sup>.

## 5. CÓMO LOGRARLO

La visión de concesiones de sistema completo para eficiencia energética no se puede lograr en un solo paso individual. Los sistemas de software actuales no están equipados en las formas descritas, ni son aplicaciones escritas de forma que permitan aprovechar esa capacidad. En términos pragmáticos, ¿cómo se puede alcanzar este resultado, y qué medidas están ya en marcha?

Como una primera consideración, los sistemas deben ser revisados para prestar atención al uso de energía; el sistema operativo en sí mismo, que siempre está en funcionamiento, todavía no ha optimizado su propio uso de energía. Hasta la fecha casi todo el software, incluyendo el software de sistemas, ha sido optimizado para rendimiento, robustez y escalabilidad pero sin considerar a la energía. Un paso inicial, por lo tanto, es rediseñar e implementar la eficiencia energética en el funcionamiento del sistema operativo. Esta es una tarea importante, y sus implicaciones todavía no se entienden completamente.

No es claro si es factible modificar los sistemas operativos existentes para que consideren a la energía como un obstáculo de primera clase, algo que sin duda sería preferible. La experiencia con sistemas de seguridad muestra que es muy complicado introducir tales consideraciones después de hechos. Con seguridad, y debido a la presión de la eficiencia energética, se pueden anticipar nuevas y fundamentales estructuras dentro del software de sistemas, e incluso que surgirán nuevos sistemas operativos. Por lo menos, se debe adaptar las facilidades de administración de recursos en el sistema operativo para la concientización y optimización energética.

*Procesadores.* Dada la importante fracción de poder que se atribuye a las CPU en las plataformas contemporánea –y como consecuencia la introducción temprana de características de administración de poder en ellos–, se ha logrado muchos avances en la programación de los sistemas operativos. La descuidada activación de hardware cuando no es

<sup>6</sup> Se reconoce que tal reduccionismo puede ser demasiado optimista si hay interacciones entre los recursos asignados por diferentes subsistemas, y que un enfoque más holístico –por ejemplo, un amplio enfoque de programación dinámica–, puede entonces ser necesario en sistemas donde "todos los joule cuentan".

útil para algún trabajo se debe eliminar. La revisión constante de un programa "director" al interior del sistema operativo –o de las aplicaciones– es un ejemplo obvio; pero éste utilizar un reloj de alta frecuencia con interrupciones de base para temporizar los eventos, manejar la hora normal, y programar la ejecución de subprocesos, algo que puede ser bastante problemático. El objetivo es mantener el hardware en reposo hasta que se necesite.

El proyecto de núcleo "tickless" [11] en Linux introduce una aplicación inicial de las marcas de aprobación dinámicas. Reprogramando el temporizador periódico de interrupciones por CPU, para eliminar las marcas de aprobación del reloj durante el reposo, se puede mejorar en un factor de 10 o más la cantidad promedio de tiempo que invierte una CPU para mantener su estado de reposo, después de cada entrada de estado inactivo. Más allá de las buenas ideas que se presentan desde Linux, el proyecto Tesla en OpenSolaris también está considerando lo que significa la transición hacia un esquema más amplio basado en eventos para el desarrollo de software en el sistema operativo.

La confluencia de características en los modernos procesadores CMT –Chip Multi threading–, CMP –Chip Multiprocessor–, y NUMA –Non-Uniform Memory Access– para sistemas multi-procesador con sockets múltiples, está generando de parte de los desarrolladores la presentación de gran cantidad de nuevos trabajos orientados a implementar la eficiencia energética [12]. Teniendo en cuenta la capacidad de alterar los niveles de rendimiento, la eficiencia energética y la introducción de CPUs multi-núcleo heterogéneos<sup>7</sup>, sólo falta añadir este componente [13,14].

**Almacenamiento.** En comparación con las CPU, el poder consumido por una unidad de disco no parece especialmente grande. Un disco típico de 3.5" y 7200 RPM consume entre 7W y 8W –sólo el 10% de lo que una típica CPU multi-núcleo consume. Aunque de mayor rendimiento, uno de 10000 RPM consumen alrededor de 14W, y uno de 15000 RPM quizás pueda utilizar alrededor de 20W; entonces, ¿cuál es la preocupación? La alarmante tasa de crecimiento relativa al almacenamiento, podría cambiar rápidamente el porcentaje de potencia total que representan los dispositivos de almacenamiento. Los factores de rendimiento y fiabilidad han dado lugar a la aplicación de husos múltiples, tan común actualmente, incluso en sistemas de escritorio para implementar una simple solución RAID. Las soluciones de almacenamiento están aumentando mucho más rápido en el centro de datos.

Los chasis para servidores de gama baja, de uso rutinario en los hogares, tienen espacio para una docena o más unidades. Por ejemplo, el rack 4U de Sun para arreglos de almacenamiento tiene capacidad para 46 unidades de 3.5". Por lo que si se utilizan unidades industriales de 10000RPM ó 15000RPM, se podría incrementar el consumo de potencia entre 1.088W y 1.6kW.

Actualmente, los subsistemas de almacenamiento obviamente están en el radar atento de la energía. Hay por lo menos dos medidas inmediatas que se pueden tomar para ayudar a mejorar el consumo de energía en los dispositivos de almacenamiento. La primera es la atención directa al uso de energía en el tradicional almacenamiento basado en disco. Parte de este trabajo lo iniciaron los vendedores de hardware de discos, quienes están empezando a introducir los estados de potencia en sus unidades de disco, y otra parte la comenzaron los desarrolladores de sistemas operativos trabajando en los actuales sistemas de archivos –como Zettabyte File System ZFS–, y en la administración de los recursos de almacenamiento. La segunda, se deriva particularmente de la reciente introducción de los grandes dispositivos de memoria Flash de bajo costo. La memoria Flash llena una importante brecha de capacidad/rendimiento entre los dispositivos de memoria principal y los discos [15,16], pero también tiene enormes ventajas de eficiencia energética sobre los medios de rotación mecánica.

**Memoria.** Debido a su relativamente bajo requerimiento de potencia –por ejemplo, 2W por DIMM–, a primera vista la memoria principal parece ser un motivo de preocupación aún menos importante que los discos. Su tamaño medio en las plataformas de hardware actual, sin embargo, la puede posicionar para crecer más rápidamente. Con el foco de los fabricantes de sistemas hardware centrado sobre todo en los niveles de rendimiento –para mantenerse al día con la correspondiente demanda de rendimiento de los CPUs multi-núcleo–, es fundamental mantener siempre el ancho de banda entre la CPU y la memoria. La consecuencia ha sido una evolución en los módulos DIMM desde un canal simple a uno doble y ahora a uno triple, en las correspondientes tecnologías SDRAM DDR, DDR2 y DDR3. Aunque las reducciones en el tamaño de los procesos de fabricación –DDR3 está en la tecnología de 50 nanómetros– ha permitido subir la frecuencia del reloj y bajar un poco la potencia por DIMM. El deseo de un rendimiento aún mayor mediante el aumento de DIMMs por canal de memoria sigue incrementando el consumo total de potencia en el sistema de memoria.

Por ejemplo, un servidor con un sistema de cuatro sockets –basado en el CPU Niagara2 de ocho núcleos de Sun–, con 16 módulos DIMM por

<sup>7</sup> Son CPUs multi-núcleo en las que los núcleos de diferentes niveles de rendimiento se colocan en el mismo paquete multi-núcleo, por lo que su consumo de poder es muy diferente.

socket y usando tecnología de memoria de doble canal DDR2, tiene en total 64 módulos DIMM; y podría incrementarse a 24 módulos DIMM por socket –96 en total–, si su sucesor más rápido utiliza en su lugar memoria de triple canal DDR3. Un DIMM DDR2 representativo consume 1.65W –ó 3.3W por par–, mientras que la edición de más baja potencia de los actuales módulos DIMM DDR3 consume 1.3W –ó 3.9W por cada tres. El resultado parece ser un incremento de sólo 20% en el consumo de potencia –en total entre 100W y 120W para nuestro ejemplo.

Sin embargo, teniendo en cuenta que la siguiente generación de CPUs también tendrá dos veces más de núcleos por zócalo, un escenario posible resultante es el deseo de tener dos veces el número de conjuntos de memoria por sockets – para un posible total de 192 módulos DIMM–, con el objetivo de equilibrar el rendimiento del sistema de memoria en general. El resultado, por lo tanto, podría ser un incremento de 100W a 240W ¡140% más de consumo de potencia para el sistema de memoria completo! Esta tendencia se observa incluso en las máquinas de clase escritorio, sin duda en una escala mucho menor, como los sistemas que han aparecido conteniendo CPUs hyperthreaded de cuatro núcleos –como el Nehalem de Intel.

Si la memoria física disponible se puede activar y desactivar, y quizás correspondientemente reconfigurar como una capacidad de procesamiento del sistema, se podrá ajustar dinámicamente algunas funcionalidades nuevas que se requerirán del subsistema de administración de memoria del sistema operativo. Mirando al futuro, se constituye en un problema abierto el diseño de un sistema de memoria virtual que sea consiente de la energía, y que sea capaz de ajustar los recursos físicos de la memoria mientras está en funcionamiento.

I/O. Los aspectos energéticos del sistema I/O en las plataformas hardware probablemente sean más importantes. Como un simple ejemplo, los actuales sistemas y subsistemas interconectados de redes de área local se han desarrollado en dos aspectos importantes: la agregación de enlaces, que cada vez es más utilizada para impulsar las redes y la fiabilidad de la banda ancha; y la velocidad de interconexión individual, que ha avanzado de 1GB a 10GB, y se espera llegue pronto a 40GB. Un transmisor-receptor para una tarjeta de interfaz de red de 10GB puede requerir hoy, como mucho, 14W cuando funciona a toda velocidad, con una consecuente reducción de potencia cuando su velocidad de vínculo se reduce a 1GB o menos –alrededor de 3W a 1GB y 1W a 100MB. Se puede esperar que otras interconexiones de alta velocidad, como InfiniBand, tengan similares consideraciones de energía para el sistema completo. Se ha prestado

poca atención a las implicaciones energéticas de las interconexiones de comunicación en cualquiera de sus diferentes manifestaciones arquitectónicas, desde un chip hasta redes de área amplia.

### 5.1 La evolución del software de aplicaciones

El aspecto más estratégico de la informática energéticamente eficiente será la evolución del software de aplicación para facilitar la eficiencia energética del sistema completo. Aunque sin duda, apoyando el desarrollo de nuevas aplicaciones de eficiencia energética, podremos esperar nuevas interfaces de aplicación para el software del sistema, la transición de las aplicaciones históricas y las actuales representa una evolución a largo plazo. Mientras tanto, ¿cómo vamos a abordar el problema de una mayor eficiencia energética para el resto de la base instalada? Obviamente, no se producirá como resultado de una implementación única de todas las aplicaciones existentes.

Una posibilidad para abordar el agnosticismo de energía de las aplicaciones existentes es realizar un análisis extrínseco de su comportamiento en tiempo de ejecución. Los datos empíricos se pueden coleccionar del grado en que el rendimiento de la aplicación es sensible a los diferentes niveles y a los tipos de aprovisionamiento de recursos. Por ejemplo, se puede observar el grado en que el rendimiento<sup>8</sup> se incrementa por la adición de recursos a la CPU, o la adjudicación de una CPU con micro-arquitectura de mayor rendimiento, y así sucesivamente [14]. La aplicación puede entonces indicar, en su forma binaria, cuál es la medida de su grado de sensibilidad, sin necesidad de alterar la implementación existente. El sistema operativo podría usar los datos para asignar los recursos que persigan cierto nivel de rendimiento especificado, o para localizar una adecuada compensación de consumo energía vs. rendimiento.

Inevitablemente, se espera que se necesite una combinación de ambas técnicas, en la que explícitamente la propia aplicación le informa al sistema de su carga de trabajo y de las necesidades de aprovisionamiento de recursos; e implícitamente en que análisis estático y dinámico se utilizan para modelar la necesidad de recursos, en relación con el rendimiento y el consumo de energía.

## 6. CONCLUSIÓN

Todavía estamos en el debut de la conciencia energética computacional; con mucha atención de parte de la industria, que ha permitido la introducción y el uso de mecanismos y controles de administración de energía en los componentes individuales de hardware, pero poca atención al

<sup>8</sup> Esto supone que se debe definir un objetivo de métrica de rendimiento externa, que puede ser problemático.

problema más amplio de la eficiencia energética: la minimización del total de energía requerida para ejecutar cargas de trabajo computacional en un sistema.

En este trabajo se propone un enfoque general a la eficiencia energética en sistemas computacionales, y se propone la implementación de mecanismos de optimización energética en el software de los sistemas, equipados con un modelo de potencia para el hardware del mismo; y se describen las aplicaciones que sugieren ajuste de aprovisionamiento de recursos para que puedan alcanzar sus niveles de carga de trabajo y/o plazos de ejecución requeridos.

En el corto plazo se requerirá, para aplicación práctica, una serie de técnicas heurísticas diseñadas para reducir los más obvios desperdicios de energía asociados con componentes de alta potencia, tales como las CPUs. A más largo plazo, y para una más efectiva optimización de la energía total, se necesitarán técnicas para modelar el rendimiento relativo a la configuración del hardware del sistema –y por tanto su consumo de energía–, junto con una mejor comprensión y un conocimiento predictivo de las cargas de trabajo.

## REFERENCIAS

1. S. Chu, "The energy problem and Lawrence Berkeley National Laboratory". Talk given to the California Air Resources Board. 2008. In: <http://www.arb.ca.gov/research/seminars/chu/chu.pdf> May 2010.
2. Environmental Protection Agency, "EPA report to Congress on server and data center energy efficiency". 2007. In: [http://www.energystar.gov/index.cfm?c=prod\\_development.server\\_efficiency\\_study](http://www.energystar.gov/index.cfm?c=prod_development.server_efficiency_study), Feb, 2010.
3. R. H. Katz, "Tech titans building boom". *IEEE Spectrum*, Vol. 46, No. 2, pp. 40-54. 2009.
4. L. Barroso & U. Holzle, "The case for energy-proportional computing". *IEEE Computer*, Vol. 40, No. 12, pp. 33-37. 2007.
5. M. Kyrnin, "PC Power Supply Efficiency. How The Efficiency Rating of a Power Supply Can Save You Money". About.com guide. 2010. In: <http://compreviews.about.com/od/cases/a/PSUEfficiency.htm>, May. 2010.
6. Energy Information Administration, "Residential Energy Consumption Surveys 2001". 2001. In: [http://www.eia.doe.gov/emeu/recs/byfuels/2001/byfuels\\_2001.html](http://www.eia.doe.gov/emeu/recs/byfuels/2001/byfuels_2001.html), May 2010.
7. K. W. Roth & K. McKenney, "Energy consumption by consumer electronics in U.S. residences". Cambridge: TIAX LLC. 2007. 147 p.
8. M. Garrett, "Powering down". *Communications of the ACM*, Vol. 51, No. 9, pp. 42-46. 2008.
9. D. J. Brown & S. C. Reams, "Toward energy-efficient computing". *Communications of the ACM*, Vol. 53, No. 3, pp. 50-58. 2010.
10. F. Yao, A. Demers & S. Shenker, "A scheduling model for reduced CPU energy". *Proceedings of 36th Annual Symposium on Foundations of Computer Science*. Milwaukee, WI, USA, pp. 374-382. 1995.
11. S. Siddha, "Getting maximum mileage out of tickles". *Proceedings of the Linux Symposium*. Ottawa, Canada, pp. 201-208. 2007.
12. A. Fedorova, "Operating system scheduling for chip multithreaded processors". Ph.D. dissertation. Harvard University, The Division Of Engineering And Applied Sciences. 2006.
13. A. Fedorova, J. C. Saez, D. Shelepov & M. Prieto, "Maximizing Power Efficiency with Asymmetric Multicore Systems". *ACM Queue Magazine*, Vol. 7, No. 10. 2009.
14. D. Shelepov, J. C. Saez, S. Jeffery, A. Fedorova, N. Perez, Z. F. Huang, S. Blagoduro & V. Kumar, "HASS: a scheduler for heterogeneous multicore systems". *ACM Operating System Review*, Vol. 43, No. 2, pp. 66-75. 2009.
15. A. Leventhal, "Flash storage today". *ACM Queue*, Vol. 6, No. 4, pp. 25-30. 2008.
16. J. Mogul, E. Argollo, M. Shah & P. Faraboschi, "Operating system support for NVM+DRAM hybrid main memory". *Proceedings of Usenix HotOS XII*. Monte Verità, Switzerland. 2009.