

Planificación de eficiencia energética en centros de supercómputo utilizando energías renovables

Energy-aware scheduling of supercomputing centers using renewable energy

Sergio Nesmachnow , Santiago Iturriaga , Universidad de la República Montevideo, Uruguay

E-mail: sergion@fing.edu.uy, siturria@fing.edu.uy

PALABRAS CLAVE:

planificación y control, supercómputo, eficiencia energética, energías renovables

KEYWORDS:

scheduling and control, supercomputing, energy aware, green energy

RESUMEN

Este artículo presenta los avances en la definición de un modelo para la planificación de eficiencia energética en centros de supercómputo considerando la utilización de energías renovables. Se plantea la gestión energética como un problema de optimización multiobjetivo y se propone una metaheurística evolutiva multiobjetivo para su resolución. Los objetivos del problema contemplan los puntos de vista de la gestión de la infraestructura computacional y la calidad de servicio ofrecida a los usuarios, considerando métricas de consumo energético mediante un perfil de potencia, valores de temperatura del sistema y calidad de servicio evaluada por la satisfacción de la fechas de finalización de trabajos enviados por los usuarios. Los resultados de la estrategia de planificación utilizando la metaheurística evolutiva multiobjetivo indican que el problema de gestión puede resolverse eficientemente, alcanzando satisfactorios niveles de costo operacional y calidad de servicio proporcionada por la infraestructura.

ABSTRACT

This article presents an energy-aware model for scheduling datacenters considering green energy. We use a multiobjective approach and we propose a multiobjective evolutionary algorithm to tackle the scheduling problem. The objectives for solving the problem include both infrastructure usage and user quality of service. We considered an energy consumption metric and temperature value for modelling infrastructure usage, and a deadline violation metric for modelling the user quality of service. The scheduling algorithm proposed in this work shows the problem can be solved efficiently reaching reasonable values for operational cost and user quality of service.

Recibido: 10 de julio de 2015 Aceptado: 10 de mayo de 2016 • Publicado en línea: 28 de febrero de 2017

1 INTRODUCCIÓN

El consumo energético de las infraestructuras de supercómputo se ha transformado en una de los principales problemas a resolver por parte de los investigadores en esta área [1]. Diversas estrategias se han propuesto para reducir los costos económicos y mitigar el impacto en el medio ambiente de la utilización de grandes cantidades de energía para la operación de los grandes sistemas de cómputo requeridos por los centros de investigación modernos. Una de las estrategias más promisorias consiste en combinar la utilización de energías renovables (por ejemplo, solar, eólica, etc.) y la utilización de técnicas de inteligencia computacional para controlar dinámicamente la operación del sistema de supercómputo de acuerdo a criterios prefijados de utilización de recursos energéticos y de calidad de servicio ofrecida a los usuarios.

En un modelo basado en control se buscan métodos automáticos eficaces para ajustar la utilización de energía y reducir los costos asociados, mediante la planificación de las tareas computacionales que son enviadas por los usuarios de la infraestructura. Las tareas pueden ser planificadas para ejecutar en los períodos en que la energía tiene un menor costo económico, en los momentos en que se dispone de mayor cantidad de energía de fuentes renovables, o siguiendo otros criterios que permitan utilizar racionalmente los recursos energéticos, siempre teniendo en cuenta las posibles degradaciones en la calidad de servicio ofrecida a los usuarios de la infraestructura.

Este trabajo presenta un modelo multiobjetivo para el problema de planificación de eficiencia energética en centros de supercómputo considerando la utilización de energías renovables, y propone un método de inteligencia computacional para la resolución del problema de control relacionado. Los objetivos considerados en la estrategia de planificación incluyen el consumo energético, el costo económico de la utilización de energía, la temperatura del sistema computacional y la calidad de servicio (en inglés, Quality of Service, QoS) proporcionada por la infraestructura.

El artículo se organiza como se describe a continuación. La Sección 2 presenta el modelo para los centros de supercómputo y la formulación del problema de control para gestión de eficiencia energética. La metaheurística multiobjetivo propuesta para la resolución del problema se describe en la Sección 3. La Sección 4 presenta el análisis experimental de los

métodos propuestos, sobre un caso de estudio de una infraestructura realista. Finalmente, las principales conclusiones de la investigación y las líneas de trabajo actual y futuro se comentan en la Sección 5.

2 DESCRIPCIÓN DEL SISTEMA DE AFORO SA-1.

2. El problema de planificación de eficiencia energética en centros de supercómputo con energías renovables

Esta sección presenta el problema de planificación de eficiencia energética abordado en este trabajo.

2.1. Modelo para los centros de supercómputo

El modelo para centros de supercómputo se basa en el previamente propuesto en nuestro trabajo [1], extendido en esta nueva propuesta para considerar la utilización de energías renovables.

En el modelo se contemplan dos componentes en la infraestructura del centro de supercómputo: el sistema de control de temperatura y enfriamiento (Heating-Ventilation-Air Conditioning, HVAC, por su sigla en inglés) y los propios servidores para cómputo que conforman la infraestructura tecnológica (IT). Estos componentes serán gestionados convenientemente para hacer un uso racional de los recursos energéticos. La Figura 1 presenta el esquema del modelo propuesto.

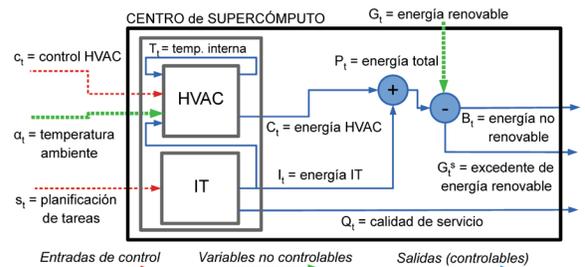


Figura 1. Modelo para gestión energética en un centro de supercómputo

En el modelo propuesto, el sistema se controla a través de las decisiones de planificación de trabajos de los usuarios a ejecutar en los servidores de la infraestructura tecnológica (determinada por una función de planificación de tareas s_t) y a través del control del subsistema de temperatura y enfriamiento HVAC. Adicionalmente, en el modelo se considera la temperatura ambiente (α_t) como una variable no controlable del sistema. La disponibilidad de energía renovable (G_t) es también un dato no controlable del problema, ya que depende de las condiciones del

sistema recolector/generador de energía renovable (energía solar, en nuestro caso).

Como salida de la gestión automatizada del sistema se tienen cuatro magnitudes: i) un valor de calidad de servicio (Q_t), que se evalúa mediante la satisfacción de fechas de finalización de los trabajos enviados por los usuarios a ejecutar en el centro de supercómputo; ii) la temperatura interna del centro de supercómputo (T_t); iii) la energía total consumida por el sistema, que es la suma de la energía utilizada para refrigeración (C_t) y la consumida por los servidores de cómputo (I_t); y iv) el total de energía no renovable utilizada (B_t) y el posible excedente de energía renovable (G_t^{\wedge}), que puede volcarse a la red eléctrica obteniendo un rédito económico.

El problema de control y planificación

La formulación del problema de control plantea la planificación de las tareas de usuario y el uso de los recursos de refrigeración durante un período de planificación, de manera que el consumo energético (P_t) no exceda un valor de referencia dado por un perfil de consumo energético prefijado (R_t), minimizando el costo de la energía no renovable y la degradación en los valores de calidad de servicio como consecuencia de tareas que pueden violar sus fechas de finalización.

El centro de supercómputo ejecuta n tareas en un período de K pasos de tiempo. Los usuarios del sistema envían tareas i para ser ejecutadas antes de una fecha límite $D(i)$, y el método de planificación determinará el momento apropiado para ejecutarlas de acuerdo a la disponibilidad de los recursos de cómputo y su consumo energético. La fecha real de finalización de la tarea i será $FT(i)$, y este valor es utilizado para evaluar la calidad de servicio de acuerdo a la satisfacción o no de la fecha límite correspondiente. El valor M_k^b define el costo monetario de la energía no renovable utilizada en cada paso de tiempo.

Formalmente, el problema multiobjetivo propone minimizar tres funciones objetivo simultáneamente:

$$O_p = \sum_{i=1}^K \begin{cases} P_k - R_k & \text{si } P_k > R_k \\ \max(R_k) & \\ 0 & \text{si } P_k \leq R_k \end{cases} \quad (1a)$$

$$O_b = \sum_{k=1}^K B_k \times M_k^b \quad (1b)$$

$$O_q = \sum_{i=1}^K \begin{cases} FT(i) - D(i) & \text{si } FT(i) > D(i) \\ 0 & \text{si } FT(i) \leq D(i) \end{cases} \quad (1c)$$

La función objetivo O_p toma en cuenta la minimización de la desviación del consumo energético con respecto al perfil de consumo utilizado como referencia. La función objetivo O_b propone minimizar el costo monetario de la energía consumida por el centro de supercómputo. Por último, la función objetivo O_q cuantifica la calidad de servicio, evaluando el tiempo total de las violaciones de fechas de finalización de tareas de usuario.

El valor de energía utilizado por el sistema de enfriamiento C_t permite controlar la temperatura global del sistema T_k , para la cual existe un límite superior tolerable para la operación correcta del centro de supercómputo. Este valor es considerado como una restricción fuerte del problema de control, por lo cual las soluciones calculadas no pueden superar el valor límite de temperatura en ningún caso.

Trabajos relacionados

Goiri et al. [2] plantearon la gestión de un centro de datos alimentado exclusivamente por energía renovable. A pesar de considerar múltiples objetivos en su modelo, la formulación del problema de optimización sigue un enfoque de objetivo único mediante un esquema de agregación de funciones, y se propone un algoritmo heurístico ávido para resolver el problema de gestión. El trabajo previo de Nesmachnow et al. [1] plantea la utilización de un enfoque multiobjetivo explícito para el problema de gestión de centros de cómputo, pero sin considerar energías renovables. Adicionalmente, se proponen heurísticas específicas basadas en backfilling para la calendarización de trabajos de usuario en los recursos computacionales.

En este trabajo se extiende el modelo previo para considerar la utilización de energías renovables, que pueden ser utilizadas para alimentar el centro de cómputo y también para obtener un rédito económico derivado de inyectar en la red los excedentes luego de aplicado el algoritmo de gestión. Adicionalmente, se propone la utilización de una técnica específica para alcanzar resultados de gran eficacia numérica al incluir una técnica de búsqueda local para potenciar a la metaheurística evolutiva multiobjetivo presentada en [1].

3 PLANIFICACIÓN DE EFICIENCIA ENERGÉTICA UTILIZANDO UNA METAHEURÍSTICA EVOLUTIVA MULTI OBJETIVO

Esta sección describe la propuesta de una metaheurística evolutiva multiobjetivo para el problema de

planificación de eficiencia energética presentado.

3.1 Metaheurísticas y algoritmos evolutivos

Las metaheurísticas son estrategias genéricas para definir algoritmos que permitan computar soluciones aproximadas de buena calidad para problemas de optimización, búsqueda y aprendizaje automático, que por su complejidad no pueden ser resueltos eficientemente por las técnicas de computación estándar [3]. La mayoría de los problemas que subyacen en situaciones del mundo real, en especial en los ámbitos científicos y tecnológicos, son lo suficientemente complejos y demandarían tiempos poco útiles en la práctica en caso de ser resueltos con técnicas clásicas de programación (enumerativas, backtracking, programación dinámica, etc.). En este contexto, las metaheurísticas emergen como métodos útiles y eficientes para resolver complejos problemas de optimización sobre escenarios reales, que requieren gran esfuerzo computacional y manejan grandes volúmenes de datos.

Los Algoritmos Evolutivos (AE) son métodos estocásticos de inteligencia computacional que emulan el proceso de evolución natural de las especies para resolver problemas de optimización, búsqueda y aprendizaje [4]. En los últimos treinta años, los AE han sido exitosamente aplicados para resolver problemas de optimización subyacentes a problemas complejos del mundo real en múltiples áreas de aplicación.

Los algoritmos evolutivos para optimización multiobjetivo (MOEA, por su sigla en inglés) [5] se utilizan para resolver eficientemente problemas del mundo real. A diferencia de los métodos tradicionales para optimización multiobjetivo, los MOEAs trabajan con un conjunto de soluciones candidatas y por ello son capaces de encontrar múltiples soluciones con diferentes valores de compromiso entre los objetivos del problema. Los MOEA deben diseñarse teniendo en cuenta dos metas: i) aproximarse a las mejores soluciones de compromiso (frente de Pareto) del problema y ii) mantener diversidad en lugar de converger a una sección específica del frente de Pareto, proporcionando diferentes alternativas para la resolución del problema.

Para mejorar la calidad de las soluciones, los AE suelen hibridarse, combinándose con otras técnicas específicas aplicadas a la resolución del problema en cuestión [6]. El nuevo algoritmo híbrido define un mecanismo de búsqueda potenciada por las buenas características de los dos algoritmos que lo componen.

En este trabajo se aplica el Non-dominated Sorting

Genetic Algorithm, version II (NSGA-II) [5], hibridado con un método de búsqueda local [7]. NSGA-II es un MOEA del estado del arte que se ha aplicado con éxito para resolver problemas complejos en múltiples ámbitos científicos y de la ingeniería. El algoritmo de búsqueda local se concibe específicamente para resolver el problema de planificación de centros de supercómputo, con el propósito de mejorar la calidad de la búsqueda de NSGA-II, tal como se describe en la próxima subsección.

3.2 Implementación

A continuación se describen las principales características de la metaheurística multiobjetivo implementada.

3.2.1 Representación de soluciones.

Cada solución representa la cantidad de energía a utilizar por el sistema de refrigeración/control de temperatura y por los servidores de cómputo en cada paso de tiempo del período de planificación, codificados en un vector de valores enteros de $2K$ elementos. Los primeros K valores, al que denominamos gen de energía del sistema de enfriamiento, indican el tipo de refrigeración utilizado: (a) un valor entre 1–100 indica que se utilizan ventiladores, y el valor indica el porcentaje de utilización de los ventiladores con respecto a su máxima velocidad, (b) un valor entre 101–200 indica que se utiliza el equipamiento de aire acondicionado, y (c) un valor entre 201–300 indica que ni los ventiladores ni el equipamiento de aire acondicionado es utilizado. Los siguientes K valores, al que denominamos gen de energía de los servidores, indican el consumo energético para los servidores, especificados en Watts.

La Figura 2 presenta un ejemplo de representación de una solución al problema de control (en azul, el gen de energía del sistema de enfriamiento; en rojo, el gen de energía de los servidores).



Fig. 2: Ejemplo de representación de soluciones, cada posición representa la energía consumida en un paso de

3.2.2 Funciones a optimizar.

Las funciones a optimizar por parte del algoritmo evolutivo corresponden a las presentadas en el modelo matemático del problema de control y planificación

Operadores para la búsqueda evolutiva en NSGA-II.

La búsqueda evolutiva de NSGA-II queda definida por los operadores de selección, recombinación y mutación en los que basan los AE.

Selección. Para la selección de soluciones para aplicar los operadores evolutivos se utiliza un operador de torneo binario, como propone tradicionalmente el algoritmo NSGA-II.

Recombinación. Para la recombinación de soluciones se utiliza un operador de cruzamiento de tres puntos, aplicado con una probabilidad p_C a cada par de individuos seleccionados, que permite combinar la información de los dos genes (energía de enfriamiento y energía de servidores) utilizados para la representación de soluciones. Se definen tres puntos de corte en la representación (puntos p_1 , p_2 y p_3); p_1 se selecciona aleatoriamente con distribución uniforme en el gen de energía del sistema de enfriamiento (posiciones 1 a K), p_2 ocupa la posición K y p_3 se define en el gen de energía de los servidores como $K+p_1$. Seleccionando los puntos de corte de esta manera, se asegura que las secciones que representan el mismo período de tiempo en ambos genes se heredan conjuntamente de padres a hijos.

Mutación. El operador de mutación se aplica con una probabilidad p_M a cada posición en la representación de soluciones. Para un valor en el gen de energía del sistema de enfriamiento (posiciones 1 a K), la mutación modifica su valor v con el valor $\text{mod}(-v + \text{rand}() \times \text{MAX_HVAC}, \text{MAX_HVAC})$, siendo MAX_HVAC el máximo valor de energía que puede consumir el sistema de enfriamiento, que corresponde a tener todas las unidades de aire acondicionado encendidas. Para los valores en el gen de energía de los servidores, se modifica su valor aleatoriamente con un valor seleccionado uniformemente entre 0 y el máximo valor de energía, que corresponde a tener todos los servidores encendidos.

Planificación de tareas.

El esquema de resolución propuesto utiliza el algoritmo NSGA-II para hallar los valores correspondientes de energía a utilizar por el sistema HVAC y por los servidores en cada paso de tiempo del período de planificación. Para determinar la asignación de recursos de cómputo a las tareas enviadas por los usuarios se utiliza una heurística específica que emplea la técnica de backfilling y un método de búsqueda local, que se describen a continuación.

Estrategia ávida para planificación utilizando backfilling. Para la planificación de tareas se adapta la heurística ávida Best Fit Hole (BFH) [8], para considerar las

características específicas del problema de planificación de eficiencia energética en centros de supercómputo utilizando energías renovables.

En la operación dinámica de un centro de supercómputo se generan periodos de tiempo libre en los servidores de la infraestructura, como consecuencia de los tiempos de arribo y ejecución de las tareas de usuario y la disponibilidad o no de recursos para ejecutarlas. Estos huecos en la planificación pueden ser aprovechados para ejecutar otras tareas, de acuerdo a la conveniencia de utilizarlos. El algoritmo BFH opera bajo la idea de aprovechar los mejores huecos disponibles para la ejecución de cada tarea de usuario. Inicialmente, BFH ordena las tareas de acuerdo a su tiempo de arribo y las asigna ordenadamente a los huecos disponibles, en caso de existir. Si una tarea puede ejecutarse en más de un hueco disponible, BFH selecciona el hueco que minimiza la diferencia entre la duración del hueco y el tiempo de ejecución de la tarea. El argumento detrás de esta decisión consiste en intentar aprovechar los huecos más justos, reservando los lugares más holgados para tareas que puedan llegar en el futuro. En el caso de que no existan huecos disponibles o no existan huecos del tamaño requerido para ejecutar una tarea, BFH selecciona para ejecutarla a la máquina que provee el mínimo tiempo de finalización para esa tarea.

Búsqueda local para mejorar la planificación de tareas. Para intentar mejorar la estrategia de planificación, se propone un método de búsqueda local específico para el problema. Los algoritmos de búsqueda local [7] han mostrado ser una de las estrategias genéricas más exitosas en la resolución de múltiples problemas de optimización, incluyendo la planificación de tareas en sistemas de cómputo [9]. A partir de una solución inicial, un algoritmo de búsqueda local busca reemplazar la solución actual por una mejor solución en una vecindad definida apropiadamente para el problema. En cada iteración se generan nuevas potenciales soluciones aplicando ciertas perturbaciones o movimientos en componentes de la solución actual, y se aceptan o descartan las nuevas soluciones generadas tomando en cuenta criterios de aceptación y rechazo.

El algoritmo de búsqueda local propuesto para el problema de planificación de tareas en el contexto de este trabajo utiliza como solución inicial para la iteración a la planificación calculada por la heurística BFH. La vecindad de una solución se define mediante un movimiento simple de una tarea desde su máquina actual a otra máquina diferente, seleccionada probabilísticamente mediante una distribución uniforme

La tarea ejecutada se inserta en una posición arbitraria, seleccionada de acuerdo a una distribución uniforme, en la máquina destino. Como criterio de aceptación de una nueva solución se utiliza la dominancia de Pareto: una nueva solución reemplaza a la actual si y solo si la nueva solución es mejor que la actual en ambos objetivos del problema.

El esquema algorítmico de resolución aplicado para el problema se basa en el concepto de metaheurísticas híbridas débiles de alto nivel [6]. El algoritmo NSGA-II se utiliza para hallar un conjunto de soluciones no dominadas del problema, sobre las cuales posteriormente se aplica el método de búsqueda local como técnica de refinamiento de las soluciones.

4. ANÁLISIS EXPERIMENTAL

Esta sección reporta los detalles del análisis experimental de la metaheurística evolutiva multiobjetivo propuesta para resolver el problema de planificación de eficiencia energética.

4.1. Instancias del problema

Las instancias del problema de control y planificación quedan definidas por un conjunto de cargas de trabajo, una configuración de hardware para el sistema computacional y un conjunto de perfiles de energía de referencia y de generación de energía renovable.

4.1.1.1. Cargas de trabajo.

En el estudio se utilizan cargas de trabajo sintéticas, generadas de acuerdo a estadísticas de utilización de centros de supercómputo [10]. Se contemplan dos tipos de cargas de trabajo: i) las correspondientes a un modo de operación normal de la infraestructura computacional, que involucran 250 tareas de usuario que ingresan al sistema en el período de planificación de 150 minutos y ii) las correspondientes a un modo de operación de alto uso de la infraestructura computacional, que involucran 500 tareas de usuario que ingresan al sistema en el período de planificación de 150 minutos.

4.1.1.2. Hardware.

La infraestructura computacional simula la del centro de cómputo Parasol, utilizado previamente en trabajos relacionados [1, 2]. Los recursos de cómputo incluyen 64 servidores con procesadores de bajo consumo energético, basados en la línea de microprocesadores Atom de Intel. Cada servidor consume 30W en el estado activo de procesamiento, 22W en estado ocioso y 3W en estado de suspensión. Considerando los 64 servidores y el sistema de enfriamiento, el centro de supercómputo puede consumir un máximo de energía de 4.2kW.

La infraestructura considerada para la generación de energía renovable consiste en un conjunto de tres paneles solares, capaces de producir un máximo de 0.5kW. La generación de energía solar se modeló considerando la información histórica utilizada en el trabajo previo [2].

4.1.1.3. Perfiles de energía.

Se consideran tres perfiles de referencia para el consumo energético, que modelan diferentes situaciones realistas en caso de disponibilidad y costo de la energía eléctrica (tarifa normal, tarifa diferenciada y tarifa escalonada por el uso de baterías y dispositivos de almacenamiento). Los perfiles de referencia para el consumo energético se presentan en la Figura 3.

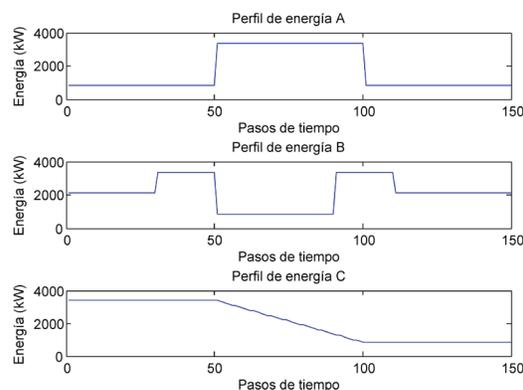


Fig. 3: Perfiles de consumo energético de referencia considerados en el estudio.

Para la generación de energía renovable se consideran tres perfiles de energía solar, correspondientes a horarios de la mañana, tarde y noche en un día típico. La Figura 4 presenta los perfiles de generación A (horario matutino) y B (horario vespertino). El perfil C asume que no hay energía renovable generada, correspondiente al horario de la noche.

4.1.1.4. Otros detalles de la simulación.

La operación del sistema computacional se simula durante un período de 150 pasos de tiempo, cada uno de un minuto de duración. La temperatura externa se considera constante a 25. La temperatura inicial del centro de supercómputo se fija en 26.5, y la máxima temperatura admisible para el centro es de 27.

Un total de 18 instancias del problema son evaluadas en el análisis experimental, contemplando las diferentes combinaciones de perfiles de energía, energía renovable y cargas de trabajo.

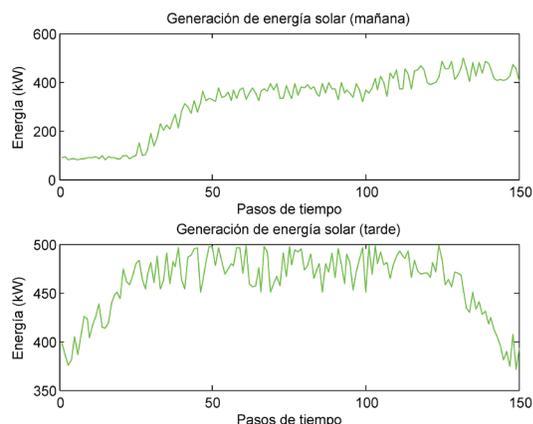


Figura 4. Perfiles de generación de energía renovable (solar)

4.2. Configuración paramétrica

Se realizó un análisis preliminar para determinar los mejores valores de los parámetros del algoritmo NSGA-II y de la búsqueda local aplicada. Los mejores resultados, considerando la calidad de las soluciones y el tiempo de ejecución de la estrategia de planificación, se obtuvieron empleando la siguiente configuración paramétrica: i) en NSGA-II, tamaño de población de 50 individuos, probabilidad de cruzamiento 0.9, probabilidad de mutación de 0.01, y criterio de parada de 500 generaciones; ii) en la búsqueda local, el criterio de parada es de 4000 iteraciones.

4.3. Resultados numéricos

A continuación se reportan los resultados obtenidos en el análisis experimental para NSGA-II y el algoritmo de búsqueda local propuesto para resolver el problema.

4.3.1.1. NSGA-II: resultados de las mejores soluciones para cada objetivo.

La Tabla 1 reporta los resultados de las mejores soluciones calculadas por NSGA-II para cada función objetivo en los casos de modo de operación normal y modo de operación de alta demanda.

Los resultados reportados en la Tabla 1 permiten comprobar que el enfoque de control evolutivo usando NSGA-II permite calcular soluciones de alta calidad para cada función objetivo. Enfocándose en las mejores soluciones calculadas, NSGA-II permite seguir el perfil de energía en hasta un 98.5% reducir el costo de operación en 57.4% al utilizar energía renovable y atender la calidad de servicio, dada por las fechas de finalización de las tareas, en un 95.0%

De acuerdo a los valores reportados en la Tabla 1, el operador del centro de supercómputo puede utilizar las soluciones de NSGA-II para enfocarse en cada uno de los objetivos del problema por separado, obteniendo muy buenos resultados, especialmente para el control y planificación de la infraestructura.

Tabla 1: Mejores soluciones calculadas por NSGA-II para cada función objetivo.

perfil de energía	ajuste al perfil de energía	reducción del costo económico	QoS
<i>Carga de trabajo: modo de operación normal</i>			
A	96.90%	50.96%	94.95%
B	97.77%	57.43%	93.70%
C	95.74%	50.27%	93.33%
<i>Carga de trabajo: modo de operación de alta demanda</i>			
A	96.00%	44.47%	88.38%
B	98.50%	49.97%	83.17%
C	86.49%	24.46%	78.75%

4.3.1.2. NSGA-II: resultados de soluciones de compromiso.

La Tabla 2 reporta los resultados promedio de las soluciones de compromiso calculadas por NSGA-II, en los casos de modo de operación normal y de alta demanda.

Tabla 2: Promedio de las soluciones calculadas con NS-

perfil de energía	ajuste al perfil de energía	reducción del costo económico	QoS
<i>Carga de trabajo: modo de operación normal</i>			
A	93.20%	41.75%	81.02%
B	94.85%	46.92%	78.05%
C	87.28%	28.23%	58.63%
<i>Carga de trabajo: modo de operación de alta demanda</i>			
A	89.47%	36.48%	42.61%
B	90.40%	44.28%	47.14%
C	70.77%	10.26%	36.96%

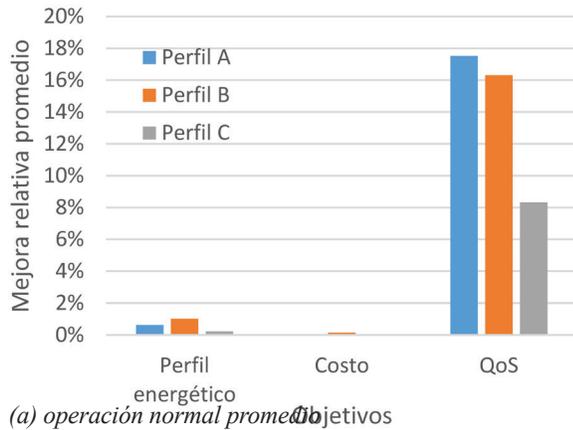
La mejor solución de compromiso se define como aquella perteneciente al frente de Pareto calculado por el algoritmo, que pondera de la misma manera los tres objetivos del problema (i.e., la solución más cercana al vector ideal del problema [5]).

Los resultados reportados en la Tabla 2 permiten indicar que las soluciones de compromiso calculadas por NSGA-II permiten seguir adecuadamente el perfil de energía predefinido para el centro de supercómputo en un 93.2% del tiempo.

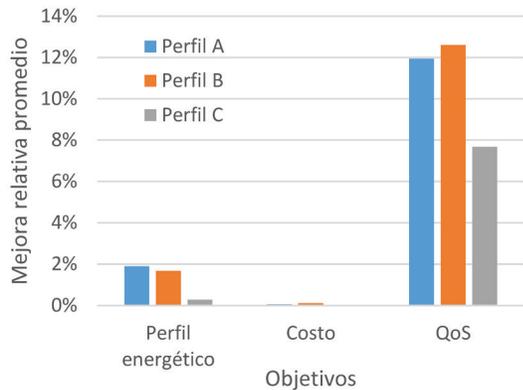
En promedio, la estrategia de control permite reducir el costo operativo en más del 30% mediante la utilización de energía renovable (en horarios diurnos, cuando la disponibilidad de energía solar está en su pico máximo). Sin embargo, los resultados también muestran que la estrategia de control se enfoca en la eficiencia energética, ya que los valores de calidad de servicio son solamente aceptables, ya que en promedio, un 60% de las tareas de usuario se cumplen dentro de la fecha de finalización establecida. Con el objetivo de mejorar este resultado de calidad de servicio fue diseñado el algoritmo de búsqueda local cuyos resultados se reportan a continuación.

4.3.1.3. Resultados de aplicar la búsqueda local.

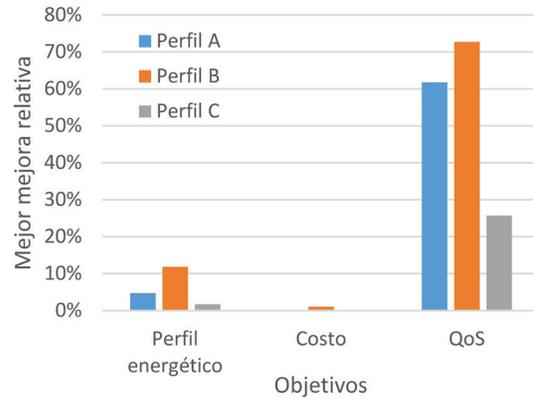
La búsqueda local se aplica sobre las soluciones calculadas por el algoritmo NSGA-II, siguiendo el modelo de metaheurísticas híbridas débiles. La Figura 5 reporta los valores promedio y máximo de la mejora obtenida al aplicar la búsqueda local, cuando se compara con los valores de las soluciones calculadas por el algoritmo NSGA-II.



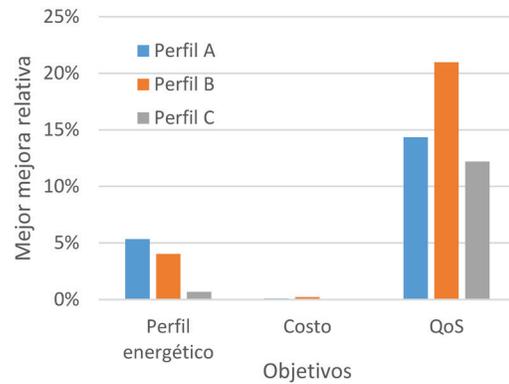
(a) operación normal promedio



(b) alta demanda (promedio)



(c) operación normal (mejor)



(d) alta demanda (mejor)

Fig. 5: Mejoras obtenidas por la búsqueda local sobre las soluciones de NSGA-II.

El análisis de los resultados reportados en la Figura 5 indica que la aplicación de la técnica de búsqueda local permite mejorar los resultados de planificación y control calculados por NSGA-II. El algoritmo de búsqueda local fue pensado para mejorar la planificación de tareas, y como consecuencia las mejoras obtenidas son significativas en el objetivo de calidad de servicio (entre 7.68–17.52% en promedio y 73.67% en el mejor caso). Este resultado es muy importante considerando que NSGA-II no muestrea apropiadamente la región del frente de Pareto que representa las soluciones del problema con mejor calidad de servicio. La búsqueda local también proporciona soluciones con mejores valores de cumplimiento del perfil de energía (0.96% en promedio y 11.80% en el mejor caso) que las calculadas por NSGA-II.

Finalmente, no se logran mejoras significativas en los costos operativos, principalmente por la buena calidad de las soluciones calculadas por NSGA-II al considerar este objetivo. Este resultado es esperado, ya que el algoritmo de búsqueda local intenta utilizar al máximo las capacidades de los recursos en estado activo, sin contemplar la posibilidad de apagar servidores.

El algoritmo de búsqueda local propuesto se enfoca apropiadamente en la calidad de servicio, proporcionando mejores soluciones y una mejor aproximación al frente de Pareto del problema. La Figura 6 presenta un ejemplo representativo de los frentes de Pareto calculados por NSGA-II y NSGA-II hibridado con búsqueda local (NSGA-II+LS), para una instancia con modo de operación normal, perfil de energía B y uso de energía renovable en horario diurno.

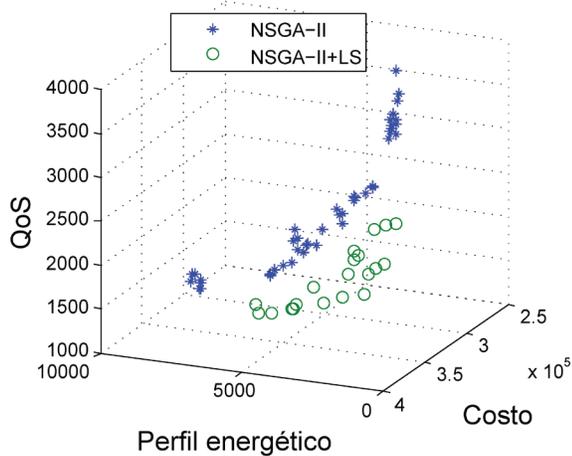


Fig. 6: Ejemplo de frentes de Pareto calculados por NSGA-II y NSGA-II+LS.

5. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo ha presentado los avances en la definición de un modelo para la planificación de eficiencia energética en centros de supercómputo considerando la utilización de energías renovables.

El problema se plantea como un problema de control y optimización, considerando múltiples objetivos a intentar optimizar simultáneamente: consumo energético, costo operativo y calidad de servicio. Se propone un enfoque de resolución basado en la aplicación de un algoritmo evolutivo multiobjetivo (NSGA-II) hibridado con un algoritmo de búsqueda local específica para el problema.

Los resultados experimentales de la evaluación sobre casos realistas del problema indican que el algoritmo NSGA-II es capaz de alcanzar buenas soluciones en cuanto a consumo energético y costo operativo del centro de supercómputo. Complementariamente, la aplicación del método de búsqueda local permite mejorar significativamente la calidad de servicio ofrecida, calculando mejores frentes de Pareto para el problema.

La utilización del algoritmo híbrido permite alcanzar valores

de hasta 98.50% de seguimiento del perfil energético de referencia (87.66% en promedio), reducir los costos operativos en hasta 57.43% al utilizar fuentes de energía renovables (34.65% en promedio) y alcanzar valores de calidad de servicio aceptables (ejecutando en el mejor caso hasta 94.95% de las tareas dentro de las fechas de finalización establecidas, y 57.41% en promedio).

Las principales líneas de trabajo actual y futuro se orientan a contemplar versiones extendidas del problema, extendiendo los horizontes de tiempo del problema de control y planificación, contemplando otros perfiles energéticos de referencia y diferentes patrones de generación de energía renovable. Asimismo, se plantea mejorar la eficiencia computacional del método propuesto, permitiendo la planificación y replanificación en línea para grandes sistemas de supercómputo.

AGRADECIMIENTOS

La investigación reportada en este trabajo ha sido parcialmente financiada por ANII y PEDECIBA, Uruguay.

REFERENCIAS

- [1] Nesmachnow, S., Perfumo, C., Goiri, I. Controlling datacenter power consumption while maintaining temperature and QoS levels. En: 3rd IEEE International Conference on Cloud Networking. 2014, 242–247.
- [2] Goiri, I., Haque, M., Le, K., Beauchea, R., Nguyen, T., Guitart, J., Torres, J., Bianchini, R. Matching renewable energy supply and demand in green datacenters. *Ad Hoc Networks*. 2015, 25(0), 520–534.
- [3] Nesmachnow, S. An overview of metaheuristics: accurate and efficient methods for optimisation. *International Journal of Metaheuristics*. 2014, 3(4), 320–347.
- [4] Bäck, T., Fogel, D., Michalewicz, Z. (eds.). *Handbook of evolutionary computation*. Oxford University Press, 1997.
- [5] Deb, K. *Multi-Objective Optimization using Evolutionary Algorithms*. New York: John Wiley & Sons, 2001.
- [6] Talbi, E.G. A taxonomy of hybrid metaheuristics. *Journal of Heuristics*. 2002, 8(5), 541–564.
- [7] Hoos, H., Stützle, T. *Stochastic Local Search: Foundations & Applications*. San Francisco: Morugan Kaufmann Publishers, Inc., 2004.
- [8] Dorronsoro, B., Nesmachnow, S., Taheri, J., Zomaya, A., Talbi, E.G., Bouvry, P. A hierarchical approach for energy-efficient scheduling of large workloads in multicore distributed systems. *Sustainable Computing*. 2014, 4(4), 252–261.
- [9] Iturriaga, S., Nesmachnow, S., Dorronsoro, B., Bouvry, P. Energy efficient scheduling in heterogeneous systems with a parallel multiobjective local search. *Computing and Informatics Journal*. 2013, 32(2), 273–294.
- [10] Nesmachnow, S. *Computación científica de alto desempeño en la Facultad de Ingeniería, Universidad de la República*. Revista de la

Acerca de los autores



Sergio Nesmachnow tiene un título de Ingeniero, de Magíster y de Doctor en Informática de la Universidad de la República, Uruguay. Actualmente es Profesor Agregado en el Centro de Cálculo, Facultad de Ingeniería, Universidad de la República. Su principal línea de investigación tiene relación con la computación científica de alto desempeño, metaheurísticas paralelas, y su aplicación para la resolución de problemas computacionalmente complejos. Ha publicado más de 50 trabajos en revistas internacionales y conferencias, es miembro del comité técnico de numerosas conferencias internacionales y es editor y revisor de prestigiosas revistas y conferencias. Email: sergion@fing.edu.uy, página web personal: <http://www.fing.edu.uy/sergion>.



Santiago Iturriaga tiene un título de Ingeniero en Computación y de Magíster en Informática de la Universidad de la República, Uruguay. Se desempeña como Docente Asistente en la Facultad de Ingeniería, Universidad de la República. Su tesis de Maestría en Informática se enfocó en la aplicación de metaheurísticas de alto desempeño para la planificación de tareas en grandes centros de cómputo. Email: siturria@fing.edu.uy, página web personal: <http://www.fing.edu.uy/siturria>.