

空調機の電力消費量を考慮したサーバ仮想化と ネットワーク動的構成変更によるデータセンタ省電力運用管理手順の提案

荒井 大輔

吉原 貴仁

株式会社 KDDI 研究所

1. はじめに

近年のトラフィック量増大に伴い、IT 機器の数も増加し、その電力消費量が 2025 年には 2006 年比で 5 倍になると推計[1]されている。そのため、IT 機器の電力消費量の削減は、トラフィック量の増大を支える上で、重要な課題である。IT 機器全体の電力消費量のうち、とくに、ルータやスイッチなどのネットワーク機器(以下、ネットワークを NW、ネットワーク機器を NW 機器と呼ぶ)とサーバが約半分を占めており[1]、それらが集中して設置されるデータセンタ(以下、DC と呼ぶ)の省電力化は急務である。

これまでに著者らは、サーバの仮想化と NW の動的構成変更による DC 省電力運用管理手順を提案している[2]。従来手順では、サーバの電力消費量にのみ注目し構成変更を実施していたため、空調機の電力消費量を考慮することで、DC 全体の電力消費量が更に省力化できる改善の余地がある。そこで、本稿では、DC 全体の電力消費量を更に省力化すること目的に、DC の電力消費をモデル化し、従来手順の拡張を新たに提案する。

2. 従来手順[2]

2.1. 従来手順の概要

多数のサーバと、その安定稼働のための空調機が稼働する DC を想定する。また、サーバは仮想化により、物理的なサーバ(以下、物理マシンと呼ぶ)の CPU やメモリ等の資源を論理分割し、1 台以上の仮想的なサーバ(以下、仮想マシンと呼ぶ)に、分割した資源を割当て、1 台の物理マシン上で複数の仮想マシンを同時実行可能とする。

従来手順では、サーバの使用率が、例えば、平日昼間は高く休日夜間は低いなど、時間帯により変動することに着目し、1 台の物理マシン上で実行する仮想マシンの数と、各仮想マシンに割当てた物理マシンの資源の量、仮想マシンを実行する物理マシンを NW セグメントの制約なく変更し、省電力化を図る。このため、省電力管理サーバを導入し、各仮想マシンに割当てた資源の使用率を定期監視する。さらに、省電力管理サーバは、監視結果に基づき、次の 2 つの手順を実行する。

(手順 1)運用者により事前指定される、物理マシンへの仮想マシンの割当て優先順に従い、1 台の物理マシン上で実行する仮想マシン数を変更する。例えば、休日夜間など、サーバの使用率が低くなる場合(図 1 左上から左下)、1 台の物理マシン上でより多くの仮想マシンを実行する。

(手順 2)仮想マシンの実行に不要な物理マシンを休止する。

2.2. 従来手順の改善点

各仮想マシンの資源の使用率と、運用者に事前指定される仮想マシンの割当て優先順により、構成変更する従来手順では、使用率が低い時間帯に仮想マシンを集約する物理マシンの組合せにより、2 つの改善点が生じる。

(改善点 1)DC 内の一部に多量の熱が発生し、その熱を除くため、空調機の電力消費量が増える。

(改善点 2)一部の多量の熱が、熱だまりの原因となる。

Proposal on Power-Saving Operation for Green Data Center Based on Server Virtualization and Network Configuration Management Counting Energy Consumption of Air Conditioner, Daisuke ARAI and Kiyohito YOSHIHARA, KDDI R&D Laboratories Inc.

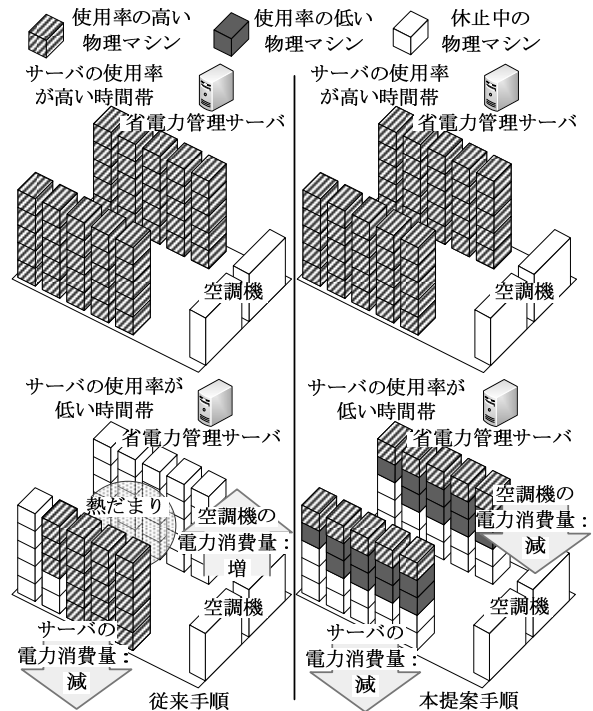


図 1: 従来手順と本提案手順の動作概要

3. データセンタの電力消費モデルの提案

空調機の冷気を床下より各サーバに供給する DC を例に、その電力消費をモデル化する。以降、3.1 節から 3.4 節まで一般的な DC における、電力消費に関わる各要素をモデル化し、3.5 節に、省電力管理サーバを採用する DC の電力消費モデルを新たに提案する。

3.1. データセンタのモデル化

空調機は、DC 内に設置され冷気を送風する CRAC (Computer Room Air Conditioner) および、DC の屋外に設置され冷気の熱源を生成する Chiller から構成される。また、サーバは、1 つのサーバラックに複数台設置される。

CRAC が送風する冷気は、DC の床下を通り、各サーバの給気面に給気される。また、各サーバが消費する電力は、そのほぼ全てが熱に変換され、各サーバの排気面より排熱される。各サーバの排気は、DC 内の天井付近を通り、CRAC と一部は再度給気面に到達する。CRAC に到達した排気は、Chiller が生成する熱源により、冷却され、再度床下より各サーバに給気される。

3.2. サーバの電力消費モデル

各サーバの電力消費(P_{serv})は、その CPU 使用率が 100% の際の電力消費(P_{max})と 0% の際の電力消費(P_{idle})、現在の CPU 使用率(n)から次式(1)により求めることができる[3]。

$$P_{serv}(n) = (P_{max} - P_{idle}) n / 100 + P_{idle} \quad (1)$$

サーバごとに、 P_{max} や P_{idle} 、 n は異なる。N 台のサーバの電力消費量の合計(P_{total})は、次式(2)で表される。

$$P_{total} = \sum_{i=1}^N P_{serv}^i(n_i) \quad (2)$$

3.3. 空調機の電力消費モデル

空調機は、CRAC と Chiller から構成される。3.1節に先述の通り、Chiller は、冷却水などの熱源を CRAC に提供するという役割から、その電力消費量は CRAC の電力消費量に密接に関係しており、本稿においては、Chiller の電力消費量は、CRAC の電力消費量に単純比例すると仮定する。

CRAC の電力消費量は、送風する冷気の温度(T_{sup})と冷却する熱、CRAC の効率(CoP, Coefficient of Performance)により決まる[4]。次式(3)に、実験的に求めた CRAC の CoP の一例[4]を示す。

$$CoP(T_{sup}) = 0.0068 T_{sup}^2 + 0.0008 T_{sup} + 0.458 \quad (3)$$

CRAC の電力消費量(P_{CRAC})は、 P_{total} と T_{sup} より、次式(4)により求めることができる[4]。

$$P_{CRAC}(P_{total}, T_{sup}) = P_{total} / CoP(T_{sup}) \quad (4)$$

式(3)および式(4)より、CRAC が送風する冷気の温度 T_{sup} が高ければ高いほど、CoP が大きくなり、空調機の電力消費量は小さくなるのが分かる。

CRAC が送風する冷気の温度 T_{sup} は、サーバの給気面の温度(T_{in})が、各サーバによりあらかじめ決まる動作保障温度($T_{redline}$)を下回るよう設定する必要がある。次に、DC 内の空気の流れより T_{sup} と T_{in} の関係を示す。

3.4. 空調機の冷気の温度設定モデル

各サーバの排気は、DC 内の天井付近を通り、CRAC と一部は再度給気面に到達する。サーバ i の排気の一部が、サーバ j の給気に混ざるサーバ j の給気量に対する割合を α_{ij} (α は 0 から 1 の実数)とする。例えば、サーバ j の給気量のうち、30%がサーバ i の排気の一部であるとする、 $\alpha_{ij}=0.3$ となる。

N 台のサーバからなる DC において、サーバ i の空気の流量と空気の密度、空気の比熱の 3 つより求まる係数(K)と、CRAC が送風する冷気の温度 T_{sup} 、サーバの給気面の温度 T_{in} 、排気面の温度 T_{out} の関係は式(5)となる[4]。

式(5)より、サーバ i の給気面の温度 T_{in} は、他のサーバからの排気(式(5)右辺第一項)と空調機からの冷気(式(5)右辺第二項)の和からなることがわかる。

$$K_i T_{in}^i = \sum_{j=1}^N \alpha_{ji} K_j T_{out}^j + (K_i - \sum_{j=1}^N \alpha_{ji} K_j) T_{sup} \quad (5)$$

3.5. サーバの休止を考慮した DC 全体の電力消費モデル

省電力管理サーバを導入した場合の DC 全体の電力消費を新たにモデル化する。省電力管理サーバは、1 台の物理マシン上で実行する仮想マシンの数と、各仮想マシンに割当てる物理マシンの資源の量、各仮想マシンを実行する物理マシンを変更可能である。そのため、省電力管理サーバは、次の 2 つを任意に変更する機能を有する。

- (1) 各物理マシンの CPU 使用率
- (2) 休止する物理マシン

省電力管理サーバが構成変更した後の各物理マシンの CPU 使用率を n_{eco} 、物理マシンの休止状態を b_{eco} (休止状態を $b_{eco}=0$ 、稼働状態を $b_{eco}=1$)とする場合、各物理マシンの電力消費量を式(6)、DC 内の温度の関係を式(7)、DC 全体の電力消費量(P_{DC})を式(8)とする。

$$P_{serv}^i(n_{eco}^i, b_{eco}^i) = ((P_{max}^i - P_{idle}^i) n_{eco}^i / 100 + P_{idle}^i) b_{eco}^i \quad (6)$$

$$K_i T_{in}^i = \sum_{j=1}^N \alpha_{ji} K_j (T_{in}^j + K_j P_{serv}^j(n_{eco}^j, b_{eco}^j)) + (K_i - \sum_{j=1}^N \alpha_{ji} K_j) T_{sup} \quad (7)$$

$$P_{DC} = (1 + 1 / CoP(T_{sup})) \sum_{i=1}^N P_{serv}^i(n_{eco}^i, b_{eco}^i) \quad (8)$$

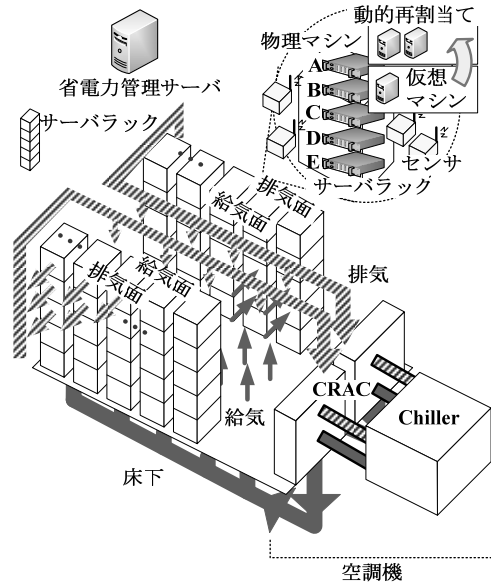


図2: データセンタの概要

4. 空調機の電力消費量を考慮した省電力運用管理手順の提案

事前準備

DC 内に温度センサを分散配置する。

運用者は、(1)物理マシン毎の電力消費モデル、(2) α_{ij} の2つを省電力管理サーバに登録する。なお、 α_{ij} はシミュレーションや DC 内に分散配置された温度センサ等により、事前または随時、得る。

省電力運用管理手順

本提案手順において、省電力管理サーバは、各仮想マシンに割当てた資源の使用率を定期監視する。省電力管理サーバは、監視結果に基づき、各物理マシンの CPU 使用率 n_{eco} と物理マシンの休止状態 b_{eco} の組を複数生成し、構成変更候補とする。次に、生成した全ての構成変更候補の DC 全体の電力消費量を求める。具体的には、式(7)にて、全ての稼働状態($b_{eco}=1$)のサーバの給気面の温度が、動作保障温度($T_{redline}$)を下回る、CRAC が送風する冷気の温度(T_{sup})を求める。さらに、式(8)より DC 全体の電力消費量を求める。最後に、最も電力消費量の少ない構成変更候補の構成に変更する(例えば、図1右上から右下)。

5. おわりに

本稿では、データセンタの省電力化を目的とする、サーバ仮想化とネットワーク動的構成変更による従来手順において、空調機の電力消費量を考慮することで、更なる省電力化を実現するため、新たに、データセンタの電力消費モデルを与え、モデルを用いた、省電力運用管理手順を提案した。本提案により、データセンタの更なる省電力化が期待できる。効率的な構成変更候補の生成手法および有効性評価が今後の課題である。

- [1] 経済産業省, “グリーン IT イニシアティブ,” http://www.meti.go.jp/press/20071207005/03_G_IT_ini.pdf.
- [2] 荒井他, “データセンタ省電力運用管理手法の提案,” 信学技報 IN2008-211.
- [3] The Green Grid, “Five Ways to Save Power,” White Paper, <http://www.thegreengrid.org/>.
- [4] Q. Thang, et al., “Thermal-Aware Task Scheduling for Data Centers through Minimizing Heat Recirculation,” IEEE Cluster Computing, 2007.