2010 年 6 月 21 日 情報システム基盤学専攻合同輪講発表資料

PowerNap: Eliminating Server Idle Power

著者： David Meisner, Brian Gold, Thomas F.Wenich
出典：*Proc. 14th Architectural Support for Programming Languages and Operating Systems, pp.205-216, Feb 2009.*
発表者： 高性能コンピューティング学講座　本多・近藤研究室 1053012　佐藤友貴

1 概要

サーバのエネルギーの大部分はサーバシステムが何もしていない状態（アイドル状態）によって浪費されている。サーバの消費電力のうち処理に利用されているのは全体の 30 %以下でそれ以外はアイドル時に消費されている。、アイドル状態のサーバはピーク消費電力の 60 %も占めている [1]。アイドル状態が続く長さ（アイドル期間）は、1 回につきわずか 1 秒程度の期間であるが、それが頻繁に起こるため、サーバの省電力化を難しくしている原因となっている。

本論文では PowerNap という省電力化手法を提案する。PowerNap はサーバの瞬間的な負荷変動に対応し、高性能なアクティブ状態と電力がほぼゼロに近い Nap 状態の間でサーバ全体の電力モードを素早く遷移させるものである。また PowerNap の省電力化効果をさらに引き出すために、電力変換効率を向上させる手法 RAILS (Redundant Array for Inexpensive Load Sharing) を提案する。

2 PowerNap

本論文では、まずサーバ全体を新しい仕事が到着するまで停止させ、低電力状態 (Nap 状態) に素早く遷移させる電力管理法を提案する。PowerNap をサーバに適用するとき、サーバがとりうる状態はアクティブ状態か Nap 状態の 2 状態のみでありシンプルである。アクティブ状態とは性能が最大の状態であり、Nap 状態とはサーバの全システムが停止し低電力な状態をいう。PowerNap の具体的な流れとしては

1. アクティブな状態で仕事を高性能モードで処理
2. 仕事の処理が終了後、サーバのシステムを停止し、Nap 状態に遷移（PowerNap 遷移）
3. Nap 状態へ遷移後、サーバは低電力状態となり、新たな仕事がサーバに到着するまでシステムは停止
4. 新たな仕事がサーバに到着後、サーバは再びアクティブ状態へ遷移（Wake 遷移）
5. アクティブ状態へ復帰し再び処理を開始
6. 1 から 5 の流れを繰り返す

となる。スリープ状態は従来モバイル機器などでは一般的に用いられてきた。しかし常時稼働状態にあることが普通であるサーバシステムではあまり用いられていない。また、短いアイドルが頻発することから高速なアクティブ/Nap 状態の遷移が求められる。PowerNap はサーバシステム向けに高速なモード遷移を提供することを特徴としており、従来のスリープ状態とは異なるものである。

3 PowerNap と DVFS の性能の比較

PowerNap によるサーバ省電力化の可能性を評価するために、従来の省電力化手法の DVFS との比較を行った。

DVFS とは CPU の処理能力要求に応じて、動作周波数とコア電圧（CPU のコア部に供給される電圧）を動的に変更 ある。

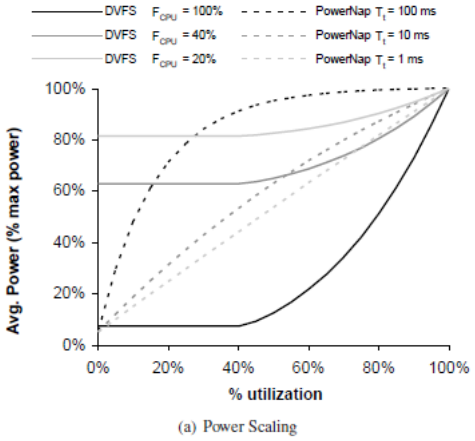

図 1: サーバの利用率と消費電力の関係

図 1 は PowerNap と DVFS におけるサーバの利用率とピーク時の電力に対する平均電力の関係を示したものである。DVFS については CPU の電力が全消費電力の何%を占めるかを F_{CPU} で表しており、PoweNap についてはモード遷移する時間を、T_t=100ms、10ms、1ms の 3 つの場合に分けて図に示している。

DVFS は利用率 40 %までは消費電力が一定でそれ以降は増加し、PowerNap は利用率に比例して消費電力が増加している。理論上では DVFS が F_{CPU}=100 %の場合が最も消費電力が低くなっているが、サーバは CPU 以外のシステムも多くあり、全消費電力を CPU が占めることは現実的にありえない。実際は CPU で消費される電力は全体の 20–40 %の間の範囲に収まる。したがって、現実的な

消費電力は F_{CPU}=20 %と F_{CPU}=40 %のグラフの間である。PowerNap の現実的な遷移時間は 10ms 程度であると考えられ、PowerNap の T_t=10ms と T_t=1ms と DVFS の F_{CPU}=20 %と F_{CPU}=40 %の間の領域を比較すると、利用率 50% 以下では PowerNap の方が消費電力が抑えられて

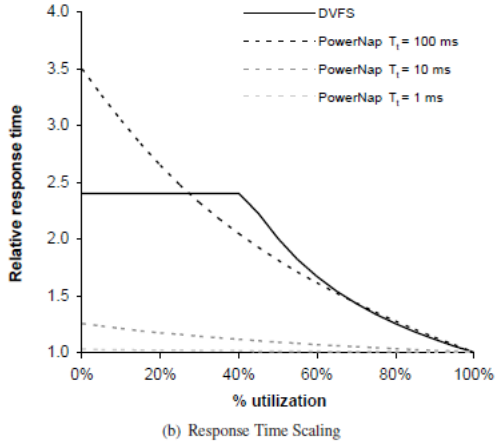

図 2: サーバの利用率と応答時間の関係

図 2 は DVFS と PowerNap のサーバ利用率と応答時間の関係を表している。応答時間とはサーバに仕事が到着してから処理が完了するまでの時間である。すべてのグラフにおいて、利用率に反比例して応答時間が短くなっているが、PowerNap の T_t=10ms と T_t=1ms では、DVFS に比べて応答時間が非常に短いことがわかる。

4 RAILS

4.1 電源変換効率

図 3 はサーバにかかる負荷と電源の電力変換効率を表したグラフである。図 3 のように電力変換効率は大きく分けて緑、黄、赤の 3 つのゾーンに分けられる。緑ゾーンは負荷 40 %以上のときで効率が 80 %から 90 %を示し、最も電力に無駄が少ない。負荷が 20–40 %では黄ゾーンとなり緑ゾーンよりも効率は低下するが電力変換効率は 70 %を超えている。しかし負荷が 20 %以下になると電力変換効率は急激に低下し、50 %以下まで低下することもある。この領域は赤ゾーンになり、電力供給に無駄が多い。

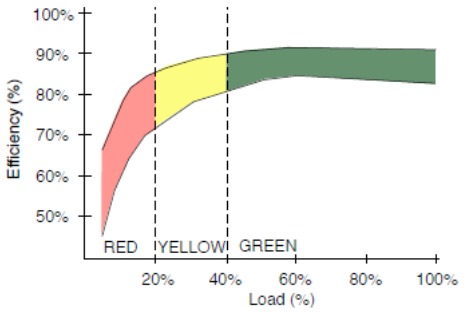

図 3: サーバに対する負荷と電力変換効率の関係

4.2 RAILS による電力変換効率の向上

PowerNap のための新しい電力供給法 RAILS (Redundant Array for Inexpensive Load Sharing) を提案する。通常、サーバ向けの電力供給法では、複数のサーバに対して高容量の電源装置 (PSU) を接続し、各サーバからの負荷に応じて電力を供給する。しかし、PowerNap を用いた場合、いくつかのサーバの電力の負荷が 0 になると PSU の負荷が小さくなり、電力変換効率が低下する。RAILS では複数のサーバに対し、容量の小さな PSU を複数個接続し、いくつかのサーバが Nap 状態のときはいくつかの PSU をオフにすることで PSU 1 台当たりの負荷を増加させ電力変換効率を向上させる。

5 RAILS の評価

電力供給法別に PowerNap を適用した場合の電力変換効率について評価した。図 4 に結果を示す。“ Commodity ”は市販の電源装置、“ 80+ ”は負荷の低い際にも高い電力変換効率を達成できる規格に適合した電源装置、“ Dynamic ”は Dynamic load-sharing を搭載した電源装置である。グラフを見ると“ Commodity ”や“ 80+ ”はサーバの利用率が 80 %以下では電力変換効率が急激に下がっているので、効率が良いとはいえない。“ RAILS ”と“ Dynamic ”については両方とも PSU のパフォーマンスが改善されている。これより RAILS は他の電力変換効率をを上回っており、PowerNap に適用するにあたっては最も適した電力供給法

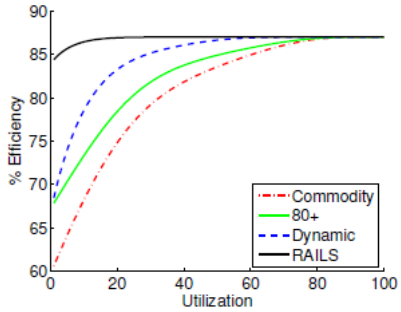

図 4: 各電力供給法の電力変換効率の比較

6 結論

本論文では PowerNap という低電力状態に素早く遷移させることによってアイドル時の電力を削減する手法を提案した。PowerNap は DVFS よりも応答時間が短く、省電力も優れていることがわかった。また、PowerNap に対して RAILS という電力供給法を用いることによって、電力変換効率が向上し、さらにサーバの省電力化を実現できることがわかった。

参考文献

[1] L. Barroso and U. H ¨ olzle, “ The case for energy-proportional computing, ” IEEE Computer, Jan 2007.