

リアルタイム電力計測に基づくストレージの省電力化手法の実装と評価

小栗 寛生[†] 引田 諭之^{††} 渡辺 陽介^{†††} 横田 治夫^{††}

[†] 東京工業大学 工学部情報工学科 〒152-8552 東京都目黒区大岡山 2-12-1

^{††} 東京工業大学 大学院情報理工学研究科 〒152-8552 東京都目黒区大岡山 2-12-1

^{†††} 東京工業大学 学術国際情報センター 〒152-8550 東京都目黒区大岡山 2-12-1

E-mail: [†]{oguri,hikida,watanabe}@de.cs.titech.ac.jp, ^{††}yokota@cs.titech.ac.jp

あらまし 近年の爆発的な情報量の増加により、ハードディスクを含むストレージシステムが大規模化し、それに伴うストレージシステムの消費電力の増加が問題となっている。そこで消費電力を抑えるためにいくつかの省電力化手法が提案されているが、既存の手法はディスクの実際の消費電力の変動等は利用しておらず、実消費電力量に基づいてディスクの構成や動作を決めるといったアプローチは採られていない。そこで本研究では、複数のディスクの消費電力をリアルタイムに測定できる環境を構築し、実際に測定したリアルタイムなディスクの消費電力の状況から最適なディスクの構成やディスクの動作を動的に決定する手法を提案し評価する。

キーワード 省電力化, ストレージ, ハードディスク

An Implementation and Evaluation of Power Saving Methods based on Real-Time Power Monitoring

Hiroki OGURI[†], Satoshi HIKIDA^{††}, Yousuke WATANABE^{†††}, and Haruo YOKOTA^{††}

[†] Department of Computer Science, School of Engineering, Tokyo Institute of Technology

^{††} Graduate School of Information Science and Engineering, Tokyo Institute of Technology

^{†††} Global Scientific Information and Computing Center, Tokyo Institute of Technology

E-mail: [†]{oguri,hikida,watanabe}@de.cs.titech.ac.jp, ^{††}yokota@cs.titech.ac.jp

Abstract Recently, power consumption of storage systems including hard disks is an important issue due to the increase of the amount of data. Although a number of methods for reducing power consumption of storage systems are proposed, these existing methods have not used actual transition of power consumption nor adjusted execution parameters of disk number or length of idle time based on measured power consumption of each disk. In this paper, we construct environment capable of real-time monitoring power consumption of each disk. We then propose methods adjusting the optimal execution parameters based on the real power consumption of each disk dynamically, and evaluate the methods in constructed environment.

Key words Energy saving, Storage, Hard disk

1. はじめに

近年の爆発的なデータ量の増加はストレージシステムを大規模化させ、それに伴いハードディスクを含むストレージシステムの消費電力の増加が問題となっている。グリーン IT 推進協議会の報告 [1] によると、日本のデータセンタの消費電力は 2025 年には 2005 年の 4 倍近くに増加すると推測されている。これは、ストレージの電力費用も増加することを示しており、ストレージの消費電力を抑えることは重要な課題になる。

ストレージの消費電力を抑えるためには、ディスクの回転を止め消費電力の低い状態を保つことが重要となる。ディスク

の状態は大きく分けてアクティブ、アイドル、スタンバイ、スリープの 4 つの状態がある [2]。アクティブは読み込み／書き込みが行われている時や起動時で、アイドルは読み込み／書き込みは行われていないがディスクが回転している状態を示している。ディスクをスピンドアウンさせ回転を停止した場合、そのディスクはスタンバイまたはスリープ状態になる。スタンバイはホストから命令があればすぐに対応できるが、スリープは多くの機能が停止しているためホストからの命令に対応するには時間がかかる。ディスクをスピンドアアップさせ回転を開始した場合は、アイドル状態に落ち着くまでに大きな電力量と時間がかかる。各状態の消費電力はアクティブが一番大きく、アイドル、

スタンバイ、スリープの順にディスクの消費電力は小さくなる。

ストレージシステムの消費電力を抑えるために、なるべく多くのディスクをスピンドアウンさせる手法には MAID(Massive Array of Idle Disks) [3] や RAPoSDA(Replica Assisted Power Saving Disk Array) [4] といったストレージシステムの省電力化手法が幾つか提案されている。これら既存の手法は、ディスクの集合を、オリジナルなデータを格納し必要に応じて回転させるデータディスクと、ディスクを常時回転させるキャッシュディスクの2つのグループに分ける。そこで、既存手法ではアクセスされたデータをキャッシュディスクに一時的に格納し、そのキャッシュディスクにアクセスを集中させることで、データディスクへのアクセスを減らしデータディスクをスピンドアウンして省電力化を実現している。

しかし既存の手法では、ディスクの動作や構成をディスクの消費電力の状況に応じて変更することは行っていない。アクセス数に対してキャッシュディスク数が大きい場合や、ほとんどアクセスされないデータディスクがある場合はディスクの動作や構成を変更することで更なる省電力効果が期待できると考えられる。

また、測定した消費電力を用いなくてディスクの動的変更を行う場合は、ディスクの仕様書などに記載されている消費電力情報を利用するしかなく、実際のディスクの状態にあった正確な変更を行うことができない。

そこで本論文では、既存の手法では利用されていないディスクの実際の消費電力の変動等を利用し、ディスクの消費電力に基づきディスクの構成や動作を決める手法を提案する。そのため本研究では、複数のディスクの消費電力をリアルタイムに測定できる環境を構築し、実際に測定したリアルタイムなディスクの消費電力の状況から最適なディスクの構成やディスクの動作を動的に決定する手法を評価する。

本論文では、2. 節で関連研究を示し、3. 節では実装したストレージシステムについて述べ、4. 節で複数ディスクの消費電力が測定できる環境について述べる。その後5. 節で実消費電力を用いた手法について述べる。6. 節では提案手法の評価実験を行い、7. 節でまとめを述べる。

2. 関連研究

本論文で前提とするストレージシステムは、複数のディスクを持ち、hdparm [5] などを用いたディスクの回転状態の制御が可能なものとする (図1 下部)。

こうした環境における既存の省電力化手法には、MAID [3] や RAPoSDA [4] などが挙げられる。MAID はディスクの集合を、オリジナルなデータを格納するデータディスクと、最近アクセスされたデータをキャッシュとして配置し常時回転させているキャッシュディスクの2グループに分ける。そして、キャッシュディスクにアクセスを集中させることで、データディスクへのアクセスを減らしアクセスの少ないデータディスクをスピンドアウンさせ省電力化を実現している。MAID は ETERNUS [6] や SATABeast [7] といった製品で採用されており実用化されている。

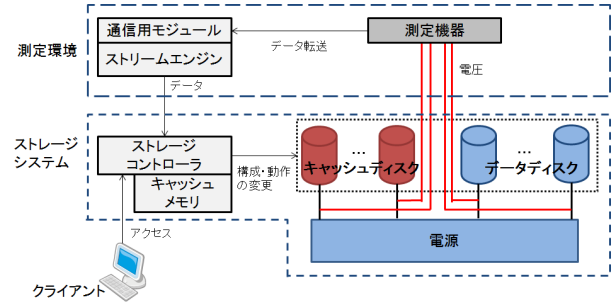


図1 リアルタイム電力計測可能な環境

RAPoSDA は MAID を拡張した手法で、データをプライマリ・バックアップ構成にすることで信頼性を向上させ、データディスクへアクセスする際にディスクの回転状況を考慮することで消費電力の抑制を行っている。

一部のディスクにデータを集める手法としては、頻繁にアクセスされるデータを一部のディスクに集めることで、一部のディスクへアクセスを偏らせアクセスが少なくなったディスクを低消費電力状態にする [8] や、アクセス関連性を用いて近い時間に使用されやすいデータを同じディスクにまとめることで不要なスピンドアアップ、スピンドアダウンを減らし省電力効果を高めている手法 [9] [10] も挙げられる。

これらの手法は、ディスクの消費電力の状況に応じてディスクの動作や構成の変更を行うことは考慮していない。そのため、アクセス数が時間によって大幅に変化し、アクセス数に対するキャッシュディスク数の割合が大きい場合や、一度アクセスされると長い時間アクセスされないデータディスクがある場合などは余計な消費電力が発生すると考えられる。

3. 実装したストレージシステム

本論文で実装したストレージシステムでは、測定した個々のディスクの消費電力情報を用いてディスクの構成や動作を動的に変更できる仕組みを用意した (図1 下部)。1つのストレージコントローラですべてのディスクを管理している。

ストレージシステム内でのデータ管理方式としてはキャッシュディスクやデータディスクを用いない通常のディスク構成の Normal と MAID を実装した。キャッシュメモリ、キャッシュディスクのキャッシュアルゴリズムは LRU とした。データの管理はキーバリュース式のデータ管理方式で行い、新規データを格納するディスクの選択方法はラウンドロビンを用いている。

3.1 ディスクの状態管理

ディスクの状態は大きく分けてアクティブ、アイドル、スタンバイ、スリープの4つの状態があり、本論文で実装したストレージシステムでは、次のようにしてディスクの状態を管理している。

ディスクへの書き込み／読み込み時はディスクの状態をアクティブ状態とし、書き込み／読み込みが終了したらアイドル状態とする。ディスクをスピンドアウンさせた場合はスタンバイ状態とし、ディスクをスピンドアアップさせた場合はアイドル状態とし

で管理している．今回実験で使用したディスク DS7K3000 [11] ではスタンバイとスリープでは消費電力の差がみられなかったため，ディスクをスピンドアウンさせることでディスクをスリープ状態にすることは行わない．

ディスクのスピンドアウンには `hdparm` を用いてディスクをスタンバイ状態にしている．スピンドアウンしているディスクに対してアクセスが来ると自動でスピンドアアップをするため，スピンドアアップはマウントしているパスに対して `ls` コマンドを実行することで行っている．

また，書き込み／読み込みの終了時刻から，同じディスクに対する次の書き込み／読み込み時までの時刻をそのディスクのアイドル時間とし，ディスクがスタンバイ状態である期間をスタンバイ時間とする．

3.2 Normal の動作

Normal ではキャッシュメモリとディスクを使いストレージシステムを構成している．

データの書き込みはライトスルーを採用し，書き込みの際はキャッシュメモリとディスクの両方に書き込む．データを読み込む場合は，キャッシュメモリ上に対象データがある場合はキャッシュメモリから読み込み，キャッシュメモリに対象データがない場合はディスクから読み込むようにする．

また，Normal ではアイドル時間が閾値を超えるディスクが発生してもスピンドアウンさせないようにしている．

3.3 MAID の動作

MAID はキャッシュメモリとキャッシュディスクとデータディスクを利用したストレージシステムを構成している．

データの書き込みはライトスルーで行い，キャッシュメモリ，キャッシュディスクの両方にデータを書き込んだ場合はデータディスクへの書き込みの終了を待たずに次の要求を処理する．キャッシュメモリ，キャッシュディスクのどちらかに書き込みなかった場合はデータディスクへの書き込みの終了を待つようにしている．データを読み込む際は，まずキャッシュメモリ上に対象データがある場合はキャッシュメモリから読み込み，キャッシュメモリになくキャッシュディスクに対象データがある場合は，キャッシュディスクから読み込み，読み込んだデータをキャッシュメモリに書き込むようにする．キャッシュメモリ上とキャッシュディスクの両方にない場合は，データディスクから読み込み，読み込んだデータをキャッシュメモリとキャッシュディスクの両方に書き込む．

アイドル時間が閾値を超えるデータディスクがある場合は該当するデータディスクをスピンドアウンさせる．

こうした MAID の動作から，ストレージシステムに対してあまりアクセスが来ない時に，キャッシュディスク数が多い場合やアイドル時間の閾値が大きい場合などは余計な消費電力が発生するため，最適なキャッシュディスク数やアイドル時間の閾値を設定することが重要になると考えられる．

4. リアルタイム電力計測可能な環境

ここでは，構築した測定環境について述べる．ディスクの消費電力を提案手法で利用するためには，複数のディスクが同時

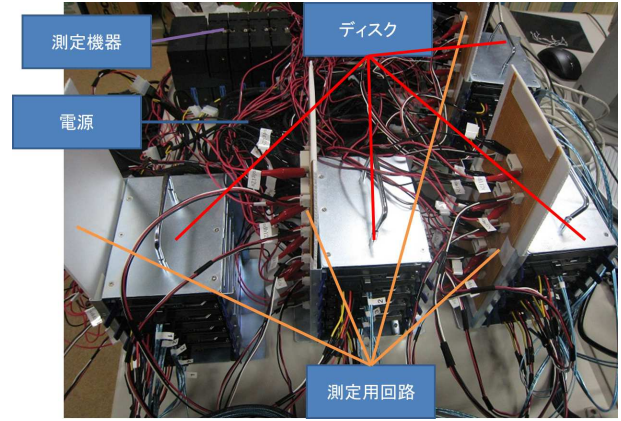


図 2 構築した測定環境

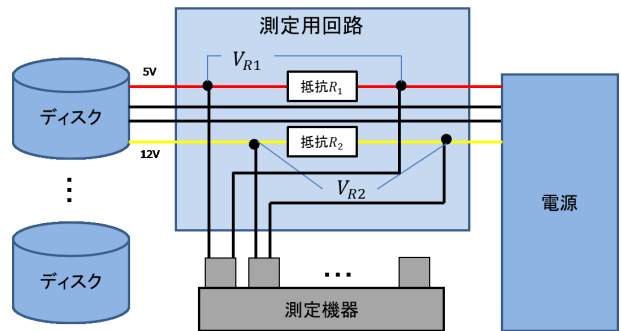


図 3 測定用回路との接続

に測定可能で，測定したデータをリアルタイムに利用できる環境が必要となる．我々は，ディスクの消費電力を測定するための回路を作成し，測定機器で測定したデータをストリームエンジンで集計することで，リアルタイムに提案手法で利用できる環境を構築した (図 1 上部，図 2)．

4.1 測定用回路

本論文では，個々のディスクの電圧を測定し，その測定した電圧から計算して各ディスクの消費電力を求めるようにする．個々のディスクの電圧を測定できるように図 3 のような回路を作成した．図 3 の測定用回路では，ディスクと電源の間に抵抗を配置し，その抵抗間の電圧降下からディスクの電圧を測定できる．ただし，抵抗値が 0.1Ω といった小さな抵抗を使用する．今回作成した回路 (図 3) の抵抗には 0.1Ω の抵抗を利用した．1 台のハードディスクに対して 5V 線，12V 線のそれぞれに抵抗を配置している．測定した電圧から消費電力の計算は，5V 線の電圧を V_5 ，12V 線の電圧を V_{12} ，5V 線に接続した抵抗の抵抗値を R_1 ，この抵抗の電圧を V_{R1} ，12V 線に接続した抵抗の抵抗値を R_2 ，この抵抗の電圧を V_{R2} として，式 1 のようにして計算して求めることができる．

$$W = V_5 \times \frac{V_{R1}}{R_1} + V_{12} \times \frac{V_{R2}}{R_2} \quad (1)$$

4.2 測定値のストリーム処理

測定したデータを提案手法で利用するには，測定した各ディスクの電圧データから消費電力の計算，計算した消費電力の平均や合計をリアルタイムに算出する仕組みが必要となる．そこで本論文では，リアルタイムなデータを処理するためにスト

MASTER stream

```
SELECT AVG(Disk1.Watt),Joule(Disk2.Watt)
FROM stream[1000]
```

図 4 StreamSpinner へのデータ要求を行うクエリ

リームエンジンを用いて各ディスクの平均や電力量を得られるようにしている。ストリームエンジンには StreamSpinner [12] を利用した。

StreamSpinner ではリアルタイムなデータに対して任意の計算を行うためのユーザ定義関数が作成できる。その関数を作成して測定した電圧から消費電力の平均や電力量を算出している。

ストレージコントローラからデータ要求を行う SQL に似たクエリを発行し、その応答を取得することで算出した結果が得られる。図 4 はクエリの例で、MASTER ではデータストリーム名、SELECT では取得したい要素、FROM では取得するデータを決めるウィンドウの長さを指定している。AVG 関数は引数で与えられた消費電力の平均を計算する。Joule 関数は引数の消費電力から電力量を計算する。このクエリではディスク 1 の 1 秒間の平均消費電力とディスク 2 の 1 秒間の電力量の要求を行っている。

4.3 測定環境の動作概要

ディスクを接続した測定用回路と測定機器を接続して個々のディスクの電圧を測定できるようになっている。そして通信用モジュールを用いて測定機器と通信し、ストリームエンジンを用いて消費電力の平均や合計を計算し、算出した結果をストレージコントローラに送信している。ストレージコントローラでは、受け取った測定結果を元に 5. 節で述べるキャッシュディスク数やアイドル時間閾値の変更を行う提案手法の処理を行う。

5. 消費電力の測定結果を用いた省電力化

既存の手法は、リアルタイムに測定したディスクの消費電力を利用しディスクの動作や構成の変更を行うことは考慮していない。

実装したストレージシステムでは、4.2 節で述べたようにストリームエンジンに対しデータ要求を行うことでディスクの消費電力が利用できるようになっている。事前に設定した時間間隔毎に消費電力や電力量の平均などを要求し取得している。取得したデータと管理しているディスクの状態を用いてディスクの構成や動作を動的に変更する。

ここでは測定した消費電力を用いたキャッシュディスク数の動的変更手法とアイドル時間閾値の動的変更手法の 2 つの手法について述べる。

5.1 キャッシュディスク数の動的変更手法 (ACD: Adjusting the number of Cache Disks)

消費電力を抑えるためには利用されないディスクをスピンドアウンさせることが効果的であり、既存の手法においてデータの利用状況やキャッシュディスク数によっては、一時的に必要なないキャッシュディスクが存在すると考えられる。そこで次の

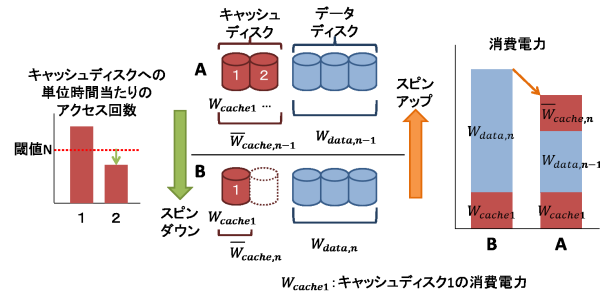


図 5 キャッシュディスク数の動的変更

ような条件を設け、キャッシュディスクをスピンドアウン、スピニアップさせることでキャッシュディスク数を変更する。

図 5 はキャッシュディスク 2 台の場合のキャッシュディスク数の動的変更手法の概略を表す図となっている。

キャッシュディスク数を減らす条件は、キャッシュディスクへのアクセス頻度が、事前に設定したアクセス頻度の閾値 F よりも少ない場合とする。ただし、キャッシュディスク数が 1 台の時はそのディスクはスピンドアウンさせない。また、キャッシュディスクをスピンドアウンさせると、残りのキャッシュディスクにアクセスが集中することやデータディスクにアクセスしに行く回数が増えることで、電力状況が大きく変化することが考えられる。そのため、一度に複数のキャッシュディスクをスピンドアウンさせることは行わない。

次に、キャッシュディスク数を増やす条件は、データディスクへのアクセスが増加しデータディスクのスピンドアウン・スピニアップが激しくなった場合に、キャッシュディスクを 1 台スピニアップさせキャッシュディスク数を増やすようにする。具体的には以下の式 2 が成り立つ場合とする。

$$W_{data,n} - W_{data,n-1} > \overline{W}_{cache,n} \quad (2)$$

ただし、 $W_{data,n}$ は n 台のキャッシュディスクをスピンドアウンさせた時のデータディスク全体の電力合計値、 $W_{data,n-1}$ は $n-1$ 台のキャッシュディスクをスピンドアウンさせた時のデータディスク全体の電力合計値、 $\overline{W}_{cache,n}$ は現在のキャッシュディスクの電力平均とする。ただし、事前に用意したキャッシュディスク数を超えてスピニアップさせることは行わない。

この式 2 を判定する周期はパラメータ $T_{interval1}$ として事前に与える。

アルゴリズム 1 はキャッシュディスク数の動的変更の疑似コードを表している。

スピンドアウンしたキャッシュディスクの扱いについて、スピンドアウン中はアクセスを受け付けないようにする。このため、スピンドアウンしているキャッシュディスクが持っているデータにアクセスがあった場合は、一度データディスクにアクセスする必要がある。データの一貫性を保つために、スピニアップした際には以前保持していたキャッシュデータは全て削除する。

5.2 アイドル時間閾値の動的変更手法 (AIT: Adjusting Idle Time threshold)

既存の手法では、データディスクのアイドル時間が事前に設定した閾値を超えた場合に、該当するデータディスクをスピ

Algorithm 1 キャッシュディスク数の動的変更

```

while true do
   $W_{data,n}, \overline{W}_{cache,n}$  を更新
  if キャッシュディスクへのアクセス頻度  $< F$  then
    キャッシュディスク数を減らす
  else  $\{W_{data,n} - W_{data,n-1} > \overline{W}_{cache,n}\}$ 
    キャッシュディスク数を増やす
  end if
  sleep( $T_{interval1}$ )
end while

```

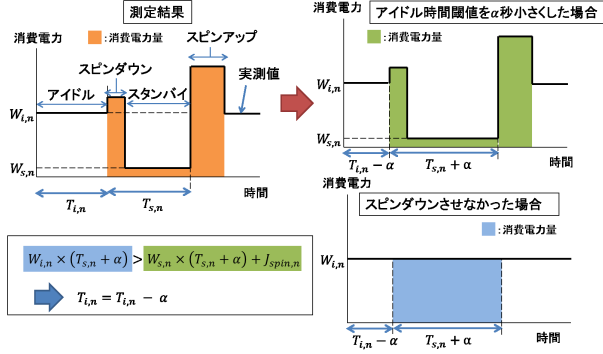


図 6 アイドル時間閾値を小さくする場合

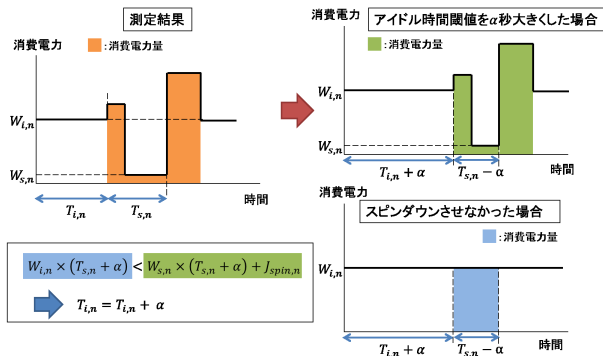


図 7 アイドル時間閾値を大きくする場合

ダウンさせている。この閾値となる時間をアイドル時間閾値と呼ぶことにする。

既存手法では、アイドル時間閾値を稼働中に変更することは考慮していない。だが、スピンアップ、スピンドウンすることが多いデータディスクがある場合は、アイドル時間閾値を大きくすることでスピンアップ、スピンドウンの回数を減らし、ほとんどアクセスされないデータディスクがある場合は、アイドル時間閾値を小さくし早めにスピンドウンさせることで、更なる省電力効果が期待できる。

図 6 と図 7 はアイドル時間閾値の動的変更手法の概略を表す図となっている。測定した結果から、アイドル時間閾値を変更した場合とスピンドウンさせなかった場合の消費電力量を比較することで、アイドル時間閾値を動的に変更する。

本研究ではストレージシステムへのアクセス要求の到着間隔が特定の分布モデルに従うことを仮定しない。 n 番目のデータディスクにおけるアイドル時間閾値 $T_{i,n}$ を事前に設定したパラメータ α を用いて、次の式 3 と式 4 がそれぞれ成立するかどうか

Algorithm 2 アイドル時間閾値の動的変更

```

while true do
   $W_{i,n}, W_{s,n}, J_{spin,n}, T_{s,n}$  を更新
  if  $W_{i,n} \times (T_{s,n} + \alpha) > W_{s,n} \times (T_{s,n} + \alpha) + J_{spin,n}$  then
     $T_{i,n} = T_{i,n} - \alpha$ 
  end if
  if  $W_{i,n} \times (T_{s,n} - \alpha) < W_{s,n} \times (T_{s,n} - \alpha) + J_{spin,n}$  then
     $T_{i,n} = T_{i,n} + \alpha$ 
  end if
  sleep( $T_{interval2}$ )
end while

```

かを条件として動的に変更する。

$$W_{i,n} \times (T_{s,n} + \alpha) > W_{s,n} \times (T_{s,n} + \alpha) + J_{spin,n} \quad (3)$$

$$W_{i,n} \times (T_{s,n} - \alpha) < W_{s,n} \times (T_{s,n} - \alpha) + J_{spin,n} \quad (4)$$

ただし、 $W_{i,n}$ 、 $W_{s,n}$ はデータディスクのアイドル時、スタンバイ時の平均電力を表し、ストレージシステムにおいて n 番目のデータディスクがアイドル状態、スタンバイ状態の時に測定した結果より算出する。 $J_{spin,n}$ はデータディスクのスピンアップ、スピンドウンに必要な電力量の和とする。スピンアップに必要な電力量はスタンバイ状態のデータディスクがアクセスされスピンアップを開始した時刻から、そのデータディスクの消費電力が測定で求めた $W_{i,n}$ になるまでの時間に消費した電力量を合計することで求める。スピンドウンに必要な電力量も同様にスピンドウンを開始した時刻から、測定で求めた $W_{s,i}$ になるまでの時間に消費した電力量を合計することで求める。また、 $T_{s,n}$ はデータディスクをスピンドウンさせた時から次のアクセスによってスピンアップが開始するまでの平均時間を表し、ストレージシステムで計算する。これらの初期値はディスクの仕様書等から得られる情報を参考にするが、十分な測定結果が得られた段階で実測値の方に切り換える。

式 3 が成立する場合はスタンバイ時間が大きく、早めにスタンバイ状態に移行した方が省電力効果が高くなると考えられる。この場合は、事前に設定したパラメータ α を用いて、アイドル時間閾値を $T_{i,n} - \alpha$ と小さくし早めにスピンドウンさせる。ただし、 $T_{i,n}$ は 0 以上とする。

式 4 が成立する場合はスタンバイ時間が小さく、ディスクをスピンドウンさせることで不要な消費電力が発生してしまう場合と考えられる。この場合は、アイドル時間閾値を $T_{i,n} + \alpha$ と大きくしスピンドウンさせにくくする。

これらの式 3、4 を判定する周期はパラメータ $T_{interval2}$ として事前に与える。

アルゴリズム 2 はアイドル時間閾値の動的変更の疑似コードを表している。

6. 実験

提案手法がストレージシステムの消費電力に対してどのような影響を与えるかを確認する実験として、実際の環境でワークロードを与えることで実験を行った。

比較する手法は、キャッシュディスクやデータディスクを用い

表 1 Normal の構成

キャッシュメモリサイズ	64MB
ディスク数	20 台
ディスクサイズ	700MB

表 2 MAID の構成

キャッシュメモリサイズ	64MB
キャッシュディスク数	2 台
キャッシュディスクサイズ	700MB
データディスク数	18 台
データディスクサイズ	700MB

表 3 実験に用いたディスク

ディスク	HDS723020BLA642
容量	2TB
キャッシュサイズ	64MB
アクティブ時の消費電力	11.3W
アイドル時の消費電力	6.8W
スタンバイ時の消費電力	0.9W
スピニングアップの電力量	450.0J
スピニングダウンの電力量	35.0J
スピニングアップ時間	15 秒
スピニングダウン時間	0.7 秒

表 4 実験に用いたワークロード

リード比率	90%
データ数	2,500
データサイズ	32KB
リクエスト数	30,000
平均リクエスト時間	100 ミリ秒
アクセス分布	Zipf
Zipf 係数	1.1
アクセス到着分布	Poisson

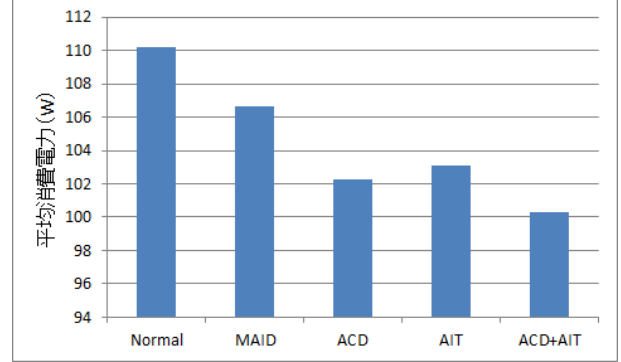


図 8 ディスク 20 台時の平均消費電力

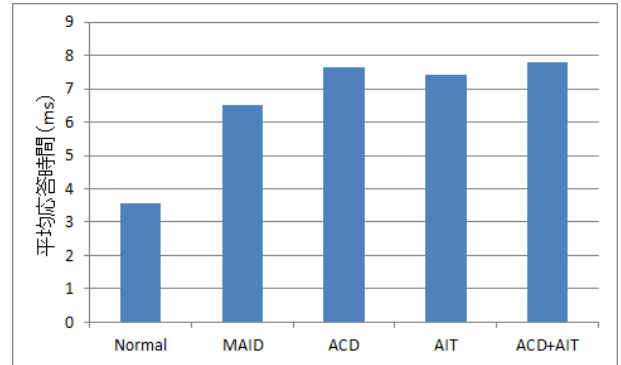


図 9 ディスク 20 台時の 1 リクエスト当たりの平均応答時間

ない通常のディスク構成の Normal と、既存手法である MAID, MAID に対してキャッシュディスク数の動的変更を適用した ACD, MAID に対してアイドル時間閾値の動的変更を適用した AIT, そして ACD と AIT を同時に行った場合の ACD+AIT で行った。

6.2 節では、それぞれの提案手法の有効性を検証する実験を行う。6.3 節では、ディスク台数の違いにより消費電力に対してどのような影響が生まれるかを確認する実験を行う。

6.1 実験条件

ディスクの電圧の測定には HIOKI のメモリハイロガー 8423 [13] を使用し、測定器によるディスクの電圧の測定間隔は 10 ミリ秒で、測定器からストリームエンジンへの測定値の送信間隔は 1 秒、ストリームエンジンからストレージシステムへの送信間隔も 1 秒としている。

Normal と MAID はそれぞれ表 1 と表 2 の構成とした。評価に用いたハードディスクは、DS7K3000 [11] で表 3 のような構成となっている。スピニングダウンの電力は文献 [11] では参考でなかったため、DS7K2000 [14] を参考にしている。評価に用いるワークロードは、人工的に生成した表 4 に示すものを利用した。キーとバリューには文字列を使用し、キーは 1 から始まる連番の文字列で、バリューは表 4 のデータサイズのものを使用した。また、提案手法の式 2, 式 3, 式 4 を判定する周期 $T_{interval1}$, $T_{interval2}$ はともに 10 秒で行った。

6.2 提案手法の消費電力と応答性能の検証実験

省電力化させていない Normal と既存研究の MAID, キャッシュディスク数の動的変更の ACD, アイドル時間閾値の動的変更の AIT, そして ACD と AIT を同時に行った場合の

ACD+AIT で比較を行う。

図 8 はワークロードを実行した際の平均消費電力, 図 9 はクライアントからリクエストを発行し、そのリクエストの結果が返ってくるまでの時間, 図 10 はワークロードの実行開始から終了までのキャッシュディスクとデータディスクのワット時を合計したものを示している。またディスク台数は 20 台, アイドル時間閾値は 30 秒, キャッシュディスクへのアクセス頻度の閾値 F は 0.6, アイドル時間閾値の動的変更に用いるパラメータ α は 5 秒としている。

図 8 の平均消費電力は、Normal が 110.1W で、MAID が 106.6W, ACD が 102.2W, AIT が 103.1W, ACD+AIT が 100.3W となった。データディスクのスピニングダウン回数は MAID が 292 回, ACD が 272 回, AIT が 336 回, ACD+AIT は 309 回となった。提案手法は、Normal と比べ約 7W から 10W, MAID と比べ約 3W から 6W 小さくすることができた。ACD の平均消費電力が AIT よりも小さくなった理由は、キャッシュディスクのスタンバイ時間がデータディスクのスタンバイ時間

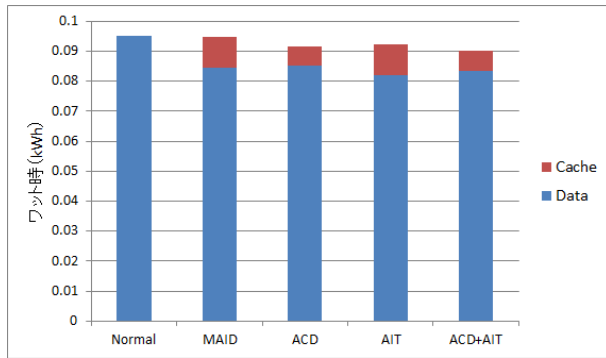


図 10 ディスク 20 台時のワット時

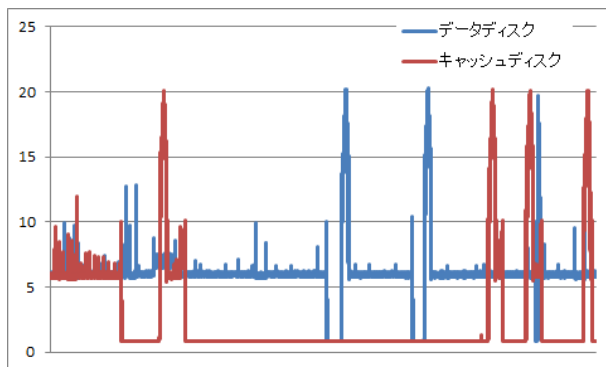


図 11 キャッシュディスクとデータディスクの消費電力の推移

と比べて非常に大きく、キャッシュディスクをスピンドウンさせることの省電力効果が大きいためであると考えられる。図 11 は、スピンドウンしたキャッシュディスク 1 台とデータディスク 1 台のワークロード実行中の消費電力の推移を表しており、縦軸は消費電力、横軸は時間を示している。この図 11 から、キャッシュディスクのスタンバイ時間の方がデータディスクよりも大きいことが確認できる。ACD+AIT の平均消費電力が最も小さくなった理由は、ACD+AIT のスピンドウン回数が AIT よりも 27 回少なくなっており、省電力効果が得られる十分なスタンバイ時間が確保できない場合を少なくすることができたためであると考えられる。

図 9 の 1 リクエスト当たりの平均応答時間は、Normal が 3.55 ミリ秒で、MAID が 6.51 ミリ秒、ACD が 7.65 ミリ秒、AIT が 7.40 ミリ秒、ACD+AIT が 7.79 ミリ秒となった。提案手法は MAID と比べ 1 ミリ秒程度の遅れで抑えることができた。提案手法が MAID と比べて遅くなった原因は、キャッシュディスクをスピンドウンさせるとキャッシュディスク上にあるキャッシュデータが削除されることとアイドル時間閾値が変更されて小さくなったことにより、スピンドウンしているデータディスクへのアクセスが多くなったためであると考えられる。

図 10 のワット時では、Normal が $0.0950kWh$ で、MAID が $0.0946kWh$ 、ACD が $0.0916kWh$ 、AIT が $0.0922kWh$ 、ACD+AIT が $0.0901kWh$ となった。Normal のワット時と比べ削減できた割合は、MAID が 0.4%、ACD が 3.5%、AIT が 2.9%、ACD+AIT が 5.2% となった。ワット時は平均消費電力と同様の結果となった。

表 5 ディスク 5 台

手法	消費電力	電力削減率	スピンドウン回数
Normal	25.9W	-	0
MAID	26.0W	-0.4%	0
ACD	22.1W	+14.7%	0
AIT	26.0W	-0.4%	0

表 6 ディスク 10 台

手法	消費電力	電力削減率	スピンドウン回数
Normal	53.4W	-	0
MAID	53.6W	-0.4%	13
ACD	50.0W	+6.4%	9
AIT	53.5W	-0.1%	14

表 7 ディスク 15 台

手法	消費電力	電力削減率	スピンドウン回数
Normal	80.2W	-	0
MAID	81.5W	-1.6%	38
ACD	76.9W	+4.2%	29
AIT	78.7W	+1.9%	38

実験より、提案手法を用いてキャッシュディスク数やアイドル時間閾値を動的に変更することで省電力効果が高まることが分かった。

また今回のワークロードでは大きなスタンバイ時間が確保しにくかったため、ワット時において提案手法は Normal と比べて顕著な差を示すことができなかったが、大きなスタンバイ時間が確保できるワークロードを使用した場合、提案手法の平均消費電力が小さいためより大きな省電力効果が期待できると考えられる。

6.3 ディスク台数が消費電力に与える影響の検証実験

次にディスク台数が提案手法に与える影響についての実験を行う。

MAID と提案手法におけるディスク構成はキャッシュディスクを 2 台とし、残りをデータディスクとしている。アイドル時間閾値は 30 秒、キャッシュディスクへのアクセス頻度の閾値 F は 0.6、アイドル時間閾値の動的変更に用いるパラメータ α は 5 秒としている。またこの実験では、ワークロードのリクエスト数を 30,000 から 5,000 に変更して行った。

比較対象は省電力化させていない Normal と既存研究の MAID、キャッシュディスク数の動的変更の ACD、アイドル時間閾値の動的変更の AIT で比較を行う。

表 5 から表 8 は、ストレージのディスク台数を 5 台、10 台、15 台、20 台と変更した場合の平均消費電力、Normal の平均消費電力と比べた削減率、スピンドウン回数を示している。

表 5 のディスク 5 台での平均消費電力は、データディスクへのアクセスが多くなってしまい、事前に設定したアイドル時間閾値以上のアイドル時間を確保できず、データディスクがスピンドウンされなかったため MAID と AIT は Normal とほぼ同じ値となった。キャッシュディスク数の割合が大きく、ACD で

表 8 ディスク 20 台

手法	消費電力	電力削減率	スピンドウン回数
Normal	112.1W	-	0
MAID	107.9W	+3.8%	53
ACD	103.7W	+7.5%	49
AIT	104.2W	+7.1%	57

はキャッシュディスク数の変更が行えたため Normal と比べて 14.7%削減できた。

表 6 のディスク 10 台での平均消費電力は、ディスク 5 台の時と同様にデータディスクへのアクセスが多くなってしまい、スピンドウンするデータディスクが少なく、MAID と AIT は Normal とほぼ同じ値となった。ACD ではキャッシュディスク数の変更が行えたため Normal と比べて 6.4%削減できた。

表 7 のディスク 15 台での平均消費電力は、AIT は Normal と比べ小さな値となったが、MAID は Normal と比べ大きな値となってしまった。MAID と AIT におけるデータディスクのスピンドウン回数が同じであることから、AIT でアイドル時間閾値を動的に変更したことにより、MAID と比べ十分なスタンバイ時間が確保できることが多くなったと考えられる。ACD では Normal と比べて 4.2%削減できた。データディスクがスピンドウンしているが、ディスク 10 台の時の Normal との差とあまり変化が見られない原因は、データディスクをスピンドウンさせた際のスタンバイ時間があまり大きくなかったことが考えられる。

表 8 のディスク 20 台での平均消費電力は、ディスク 15 台の時よりも 1 台当たりのデータディスクへのアクセスが減少し、スピンドウンさせた際のスタンバイ時間が十分確保できた場合が多かったことが考えられる。そのため MAID, ACD, AIT すべての場合の平均消費電力が Normal よりも小さな値となった。

この実験より、提案手法はディスク台数が多くなるにつれてデータディスクのスピンドウン回数やスタンバイ時間が大きくなり易く、省電力効果が得られやすいことが確認できた。

7. まとめと今後の課題

本論文では、複数のディスクの消費電力が同時に測定可能でディスクの消費電力をリアルタイムに利用できる測定環境を構築した。そして実際に測定したディスクの消費電力を利用したキャッシュディスク数の動的変更やアイドル時間閾値の動的変更を行う手法を提案し評価した。

提案手法の性能と消費電力の検証実験では、特定のワークロードでは提案手法を用いてキャッシュディスク数やアイドル時間閾値を動的に変更することの優位性を示すことができた。ディスク台数を変更した実験では、ディスク台数が多いほど提案手法は省電力効果が高くなることが確認できた。

今後の課題としては、より実行時間の長いワークロードを使用した場合の提案手法の省電力効果の検証や、リード比率やリクエスト数などが異なるワークロードを利用した場合、ディスク数などが異なる場合に省電力効果に対しどのような違いがみ

られるかを評価する必要がある。また実際のワークロードのトレースを利用した場合やキャッシュディスクを SSD に変更した場合なども課題として挙げられる。さらにディスクの回転状況を考慮した RAPoSDA についても実装し評価を行う必要がある。

謝 辞

本研究の一部は、日本学術振興会科学研究費補助金基盤研究 (A)(# 22240005) の助成により行われた。

文 献

- [1] "2009 年度 グリーン IT 推進協議会 調査分析委員会 報告書概要", [http://www.greenit-pc.jp/activity/reporting/100707/pdf/GIPC_Survey_and_Estimation_Committee_Report_FY2009_\(Summary\).jp.pdf](http://www.greenit-pc.jp/activity/reporting/100707/pdf/GIPC_Survey_and_Estimation_Committee_Report_FY2009_(Summary).jp.pdf)
- [2] 岡村 博司, 服部 正勝: 改訂 ハード・ディスク装置の構造と応用, CQ 出版社, 2010.
- [3] Dennis Colarelli and Dirk Grunwald. Massive Arrays of Idle Disks For Storage Archives. In Proceedings of the 2002 ACM/IEEE conference on Supercomputing, Supercomputing '02, pp. 1-11, Los Alamitos, CA, USA, 2002. IEEE Computer Society Press.
- [4] 引田諭之, HieuHanh Le, 横田治夫. ストレージ省電力化手法の電力削減におけるシステム構成の影響. DEIM Forum 2012 D6-2, 2012.
- [5] "hdparm." [Online]. Available: <http://sourceforge.net/projects/hdparm/>
- [6] "ETERNUS DX60 S2/DX80 S2/DX90 S2.", <http://storage-system.fujitsu.com/jp/products/diskarray/dx-entry/>
- [7] "SATABeast/SATABeastXi.", <http://www.nclc.co.jp/products/storage/nexasan/satabeast/>
- [8] Eduardo Pinheiro and Ricardo Bianchini. Energy Conservation Techniques for Disk ArrayBased Servers. In Proceedings of the 18th annual international conference on Supercomputing, pp. 68-78, New York, NY, USA, 2004. ACM.
- [9] 入谷優, 横田治夫. アクセス関連性によるストレージデータ配置の性能と消費電力. 第 152 回データベースシステム・第 103 回情報基礎とアクセス技術合同研究発表会, 2011.
- [10] 入谷優, 横田治夫. 動的に記録されたアクセス関連性に基づくデータ配置による省電力ストレージシステムの消費電力と性能に対する影響. DEIM Forum 2012 B1-4, 2012.
- [11] Technologies, H.G.S.: Hard Disk Drive Specification, Hitachi Deskstar 7k3000, [http://www.hgst.com/tech/techlib.nsf/techdocs/EF593BD721D5D2768825782D000B8111/\\$file/DS7K3000-US7K3000-SATA.OEMSpecRev1.3.pdf](http://www.hgst.com/tech/techlib.nsf/techdocs/EF593BD721D5D2768825782D000B8111/$file/DS7K3000-US7K3000-SATA.OEMSpecRev1.3.pdf)
- [12] "StreamSpinner." [Online]. Available: <http://www.streamspinner.org/>
- [13] "HIOKI メモリハイログャー 8423.", <http://hioki.jp/newproduct/8423/index.html>
- [14] Technologies, H.G.S.: Hard Disk Drive Specification, Hitachi Deskstar 7k2000, [http://www.hgst.com/tech/techlib.nsf/techdocs/5F2DC3B35EA0311386257634000284AD/\\$file/USA7K2000_DS7K2000.OEMSpec.r1.4.pdf](http://www.hgst.com/tech/techlib.nsf/techdocs/5F2DC3B35EA0311386257634000284AD/$file/USA7K2000_DS7K2000.OEMSpec.r1.4.pdf)