

# 仮想化環境における応用情報を用いたデータ再配置による ストレージ省電力化

若色匠<sup>†1</sup> 山口実靖<sup>†2</sup>

データセンターにて膨大な数の計算機が稼働しており、多くの電力が消費されている。ストレージ機器はその中でも消費電力が大きい装置の一つであり、この電力消費の削減は重要な課題の一つとなっている。ストレージ電力消費の削減手法の一つにアプリケーションの動作情報を用いてデータ配置を改善し、特定のストレージデバイスに大きなアクセス間隔を作成し省電力機能を適用させる手法がある。本研究では、当該手法を仮想化環境におけるデータインテンシブアプリケーションに適用し、ストレージ消費電力の削減を目指す。具体的には、仮想計算機上で TPC-E ベンチマークを稼働させテーブルごとのアクセス頻度を調査する。そして、特定のストレージ装置にアクセス頻度が低いテーブルを配置し、そのストレージ装置の消費電力の低減を図る。本手法を実装、評価した結果、本手法により小さな性能の低下により大幅なストレージ消費電力の削減が可能であることが確認された。

## Storage Power Saving with Applications Cooperation in Virtualized Environment

TAKUMI WAKAIRO<sup>†1</sup> SANEYASU YAMAGUCHI<sup>†2</sup>

Large power is consumed in datacenters by huge amount of computers. Storage devices are ones of power consuming devices in computers. In this paper, we propose a method for decreasing power consumption by optimizing data layout with application's supports in virtualized environment. With the method, data accesses are monitored and frequently accessed data are removed from a certain storage device in order to create large access intervals and bring the device to power saving mode. For evaluation, we executed TPC-E benchmark and measured access intervals. Then, we applied the proposed method and evaluated power consumption and performance. Our experimental results demonstrated that our method could save power consumption with small performance decline.

### 1. はじめに

データセンターにて膨大な数の計算機が稼働しており、多くの電力が消費されている。ストレージ機器はその中でも消費電力が大きい装置の一つであり、この電力消費の削減は重要な課題の一つとなっている。この問題に対する解決策の一つとして、アプリケーションの動作情報を用いてディスク上のデータレイアウトを変更し、HDD の消費電力を削減する手法がある[1]。

本研究では仮想化環境下での TPC-E の各テーブルデータから HDD へのアクセス頻度を調査した。そして、長いアクセス間隔が多く、短いアクセス間隔の少ないテーブルデータを1つの HDD にまとめ、HDD 停止時間がどの程度確保できるか、スループットの低下はどの程度起こるか調査した。また HDD において停止時間を設定し、停止によるスループットの低下はどの程度か、停止によって生じる消費電力減少の程度を調査した。

### 2. 既存研究

前章で説明した手法の一つに応用(アプリケーション)の動作情報を用いたストレージ省電力手法の一つとしてデー

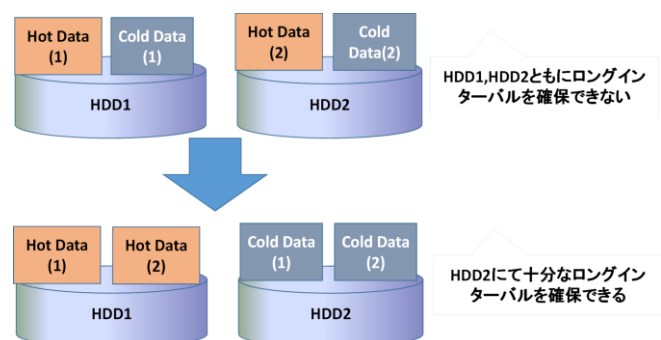


図 1 応用情報を用いたデータレイアウトの変更

タ(テーブル)のアクセス頻度を考慮し、ディスクへのデータ配置を制御することにより、ディスクの省電力機能を適用できるだけの I/O 発行間隔を生成する手法[1]が提案されている。本研究ではアクセス数が多いデータを Hot データ、アクセス数が少ないデータを Cold データと呼び、この Cold データをひとつの HDD に集中させることでアクセス間隔の拡大させ省電力化をはかる [2]。図 1 に応用情報を用いたデータレイアウトの変更法について示す。

図 2 にてストレージの停止と再起動時の電力の変化について示す。ストレージ停止により削減できる電力量とストレージ再稼働により失われる電力量が等しくなるストレージ停止時間(A=B)をブレイクイーブンタイムと呼び、それ

<sup>†1</sup> 工学院大学大学院工学研究科電気・電子工学専攻  
Electrical Engineering and Electronics, Kogakuin University Graduate School  
<sup>†2</sup> 工学院大学工学部情報通信工学科  
Department of information and Communications Engineering, Kogakuin University

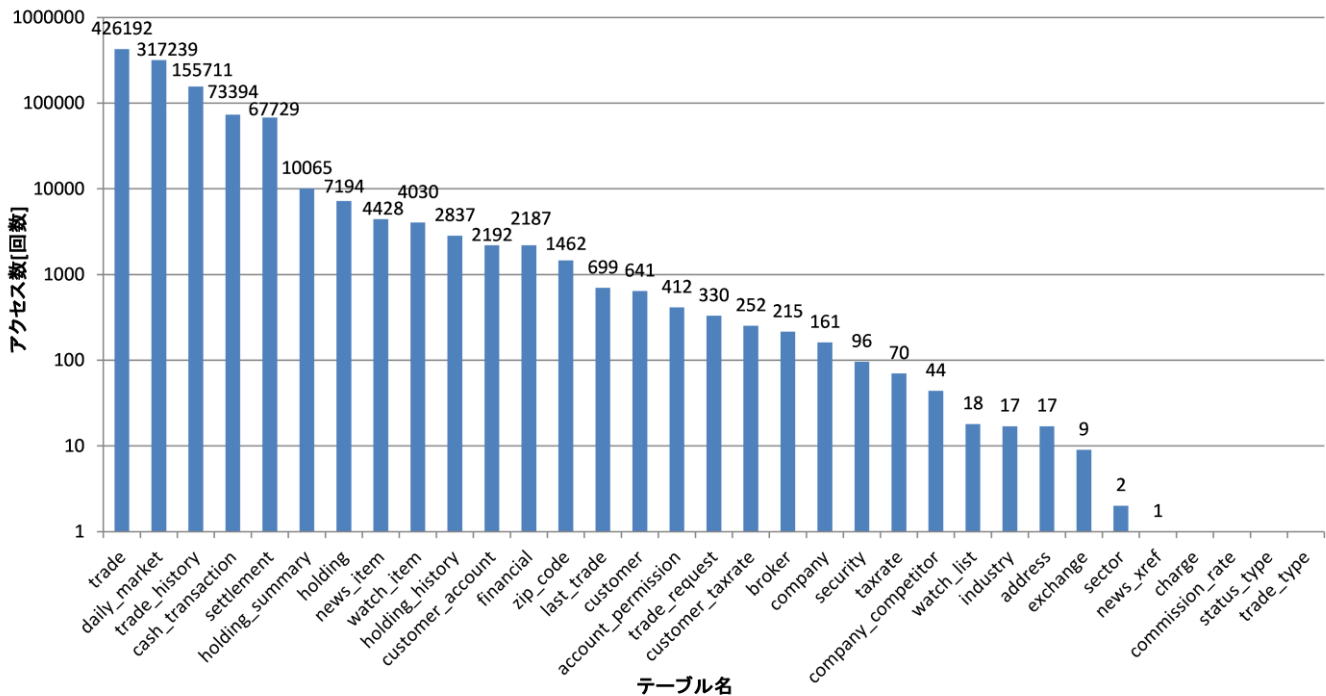


図3 各テーブルデータへのアクセス数

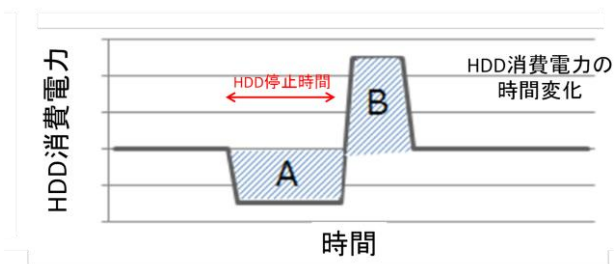


図2 停止と再起動時の電力の変化

より長くなる HDD アクセス間隔(A>B)をロングインターバルと呼び、上記手法によりロングインターバルを作り出すことで、省電力化を実現している。

文献[2,3]ではそれぞれブレイクオープンタイムが25秒、10秒と定義されており、ブレイクオープンタイムに使用したHDDと考えられる。

### 3. 基本調査

#### 3.1 アクセス頻度調査

基本調査として1台の物理計算機上に1台のVMを稼働させ、VM上でベンチマークソフトtpcmysqlを2時間実行させ、各テーブルデータへのアクセス数とアクセス頻度を調査した。調査は表1の環境にて行った。アクセス頻度はLinuxカーネルのSCSIサブシステム内で観察した。よってページキャッシュなどによりストレージアクセスを発生

(およびそれに伴う電力消費)させないアクセスは計測に含まれていない。ただし、データサイズの合計はゲストOSメモリの約16倍であり、ページキャッシュはほぼヒットしないようになっている。各テーブルデータへのアクセス数を図3に示す。図3よりテーブルデータへのアクセス数には大きな偏りがあり、tradeなどの多いものでは40万以上、exchangeなどの少ないものでは10回以下であり1度もアクセスの無いchargeなどのテーブルデータもあった。各テーブルデータへのアクセス頻度を表2に示す(表2にはアクセス数が1以下のテーブルはアクセス間隔が存在しないため省いてある)。表2より、30秒以下のアクセス間隔が10回以下であるcompany\_competitorよりアクセス数の少ないテーブルはVM数を増やしたとしても短い間隔でのアクセスが増えないと考えられるので、ロングインターバルを確保できると考えられる。また、100秒以上の間隔が比較的

表1 提案手法評価の測定環境

CPU	Intel Celeron CPU G1101 2.27[GHz]
MEMORY	4[GB]
HDD(OS用)	VB0160EAVEQ 160[GB]
HDD(データ用)	WD5000AZRX-0 500[GB]
OS(共通)	CentOS6.3 x86_64
カーネル(共通)	2.6.32.57
ホストOSメモリ	2[GB]
ゲストOSメモリ	512[MB]
仮想HDD	100[GB]
合計データサイズ	7.8[GB]

大きく見られ、30秒以下の回数が少ない company, security, taxrate もロングインターバルを得られるのではないかと考えられる。逆に、10秒以下の間隔しかない watch\_item より上のテーブルデータは短いアクセスが多くロングインターバルが存在しないため、これらのテーブルが1つでも存在すると HDD の停止時間を得ることが不可能であることが分かる。また、30秒以上の間隔がある broker や trade\_request より上のテーブルは100秒以上の間隔が無く30秒以下のアクセスも多く見られるため、ロングインターバルを得ることは難しいと考えられる。

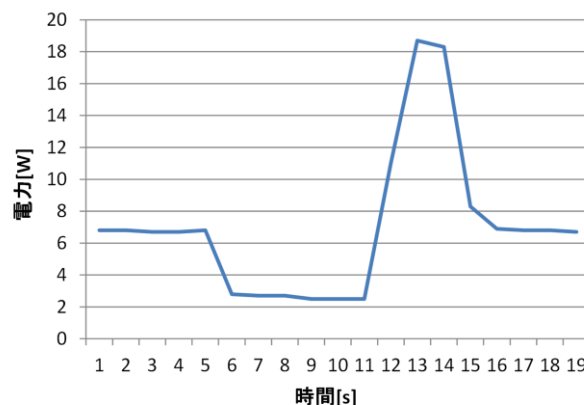


図4 スピンドアウンからスピニアップまでの推移

表2 各テーブルデータのアクセス間隔

テーブル名 \ アクセス間隔[秒]	0~10	10~30	30~100	100~
trade	426191	0	0	0
daily_market	317238	0	0	0
trade_history	155710	0	0	0
cash_transaction	73393	0	0	0
settlement	67728	0	0	0
holding_summary	9978	86	0	0
holding	6973	220	0	0
news_item	4372	55	0	0
watch_item	3974	55	0	0
holding_history	2570	233	33	0
customer_account	2031	93	67	0
financial	1980	203	3	0
zip_code	1202	246	13	0
last_trade	464	134	100	0
customer	397	185	58	0
account_permission	209	116	86	0
trade_request	87	152	90	0
customer_taxrate	180	17	26	28
broker	1	105	108	0
company	68	20	49	23
security	31	9	29	26
taxrate	16	4	16	33
company_competitor	1	1	12	29
watch_list	4	1	1	11
industry	0	0	0	16
address	2	2	3	9
exchange	0	0	0	8
sector	1	0	0	0

表3 アクセス頻度調査の測定環境

CPU	Intel Celeron CPU G1101 2.27[GHz]
MEMORY	4[GB]
HDD(OS用)	VB0160EAVEQ 160[GB]
HDD(データ用)	WD5000AZRX-0 500[GB]
OS(共通)	CentOS6.3 x86_64
カーネル(共通)	2.6.32.57
ホストメモリ	2[GB]
ゲストメモリ	512[MB]
仮想HDD	100[GB]
合計データサイズ	7.8[GB]

### 3.2 ロングインターバルの調査

HDD をスピンドアウンすると一時的に消費電力を下げることはできるが、スピニアップ時に一定の大きな電力消費が発生する。スピニアップ時に減少した電力がスピニアップ時に増加した電力より多くなる時間がロングインターバルである。測定は表3の環境で行った。アクセス頻度の測定で用いた“WD5000AZRX-0”は、hdparm コマンドによるスピンドアウンの設定をすることができなかつたため、HDDに“VB0160EAVEQ”を用いて行う。図4は今回使用したHDDのスピンドアウンからスピニアップまでの電力をワットモニターを用いて調査した結果である。図4より、5秒の位置でスピンドアウンが行われたことで消費電力が半分以下になったが、11秒の位置でのスピニアップにより一時的に倍以上の消費電力が発生していることがわかる。スピニアップの際に生じる消費電力は一定であり、このHDDでは10秒以上の停止時間がロングインターバルとなる。

## 4. データベーステーブル再配置手法

### 4.1 再配置手法

上記のようにTPC-Eではデータベーステーブルが複数作られ、テーブルのアクセス頻度には偏りがある。また、アクセス頻度もテーブルによって大きく異なり、ロングインターバルの存在するテーブルも複数存在する。本手法ではロングインターバルのあるテーブルを1つのHDDにまとめることにより、HDDのアクセス間隔を拡大し、HDDのロングインターバルを確保することにより省電力化を実現する。

### 4.2 性能評価

提案手法の有効性を確認するために性能評価実験を行った。1台の物理計算機上に3台のVMを稼働させ、それぞれのVM上でベンチマークソフトtpccmysqlを実行し、テーブルデータ移動前後の各VMのスループットとロングインターバルのあるテーブルデータを集約したHDDのアクセス間隔を調査した。測定環境は表3の通りである。ま

た、ロングインターバルのあるテーブルデータを集約した HDD にスピンドウン時間を設定し停止した場合のスループットと消費電力を調査した。

実験環境には物理 HDD が 4 台あり、1 台は OS のシステムファイルを格納し、残りの 3 台には TPC-E のテーブルデータを配置する。移動前(手法適用前)は各テーブルデータ配置用 HDD に各 VM のテーブルデータを配置する(1VM が 1HDD に対応)。移動後(適用後)はロングインターバルのあるテーブルデータを特定の HDD にまとめて配置し、その HDD 内のロングインターバルの無いテーブルデータは残りの 2 つの HDD にアクセス頻度が均等になるように配置する。今回は確実にロングインターバルが得られると考えられる図 3 の company\_competitor 以下のテーブルデータを移動し測定を行った。

ロングインターバルのあるテーブルデータを配置した HDD における移動前後のアクセス間隔の発生頻度は表 4 に示す。表 4 より、移動前ではロングインターバルは 1 回も存在しなく、停止できる時間は無かった。しかし、移動後はロングインターバルのある表のみを集めたためアクセス頻度が千分の一ほどに減少し、100 秒以上の間隔も多く得ることができた。そのため、HDD のスピンドウン設定により省電力化ができると考えられる。

各 VM 上の tpccmysql の平均スループットを図 5 に示す。図 5 内の“移動後”は再配置を行ったが HDD 停止設定は行っていない状態、“スピンドウン設定後”は再配置を行いつつスピンドウン時間を 10 秒とした状態である。vm1, vm2 共にアクセスの少ないデータを vm3 用 HDD に移動したため、この 2 つの VM のスループットは移動前後で近い値になっている。しかし vm3 はアクセスの多いデータを vm1, vm2 用 HDD に分配配置したため、移動後は vm1, vm2 よりスループットは悪くなっている。そして、3 つの HDD から 2 つの HDD にアクセスの多いファイルが集中配置されたため、移動後は各 VM ともスループットが低下しているが、低下幅は平均で約 15% となっている。また、スピンドウン時間を設定することによるスループットの変化はほと

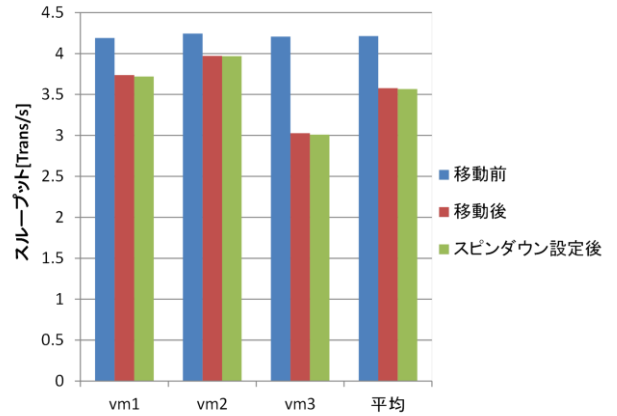


図 5 各 VM におけるスループット

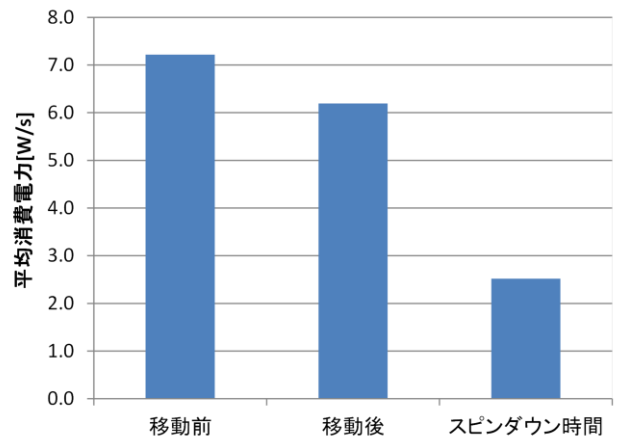


図 6 時間あたりの消費電力

表 4 移動前後の HDD アクセス頻度分布

間隔 [秒]	移動前 [回]	移動後 [回]
0~1	597665	369
1~2	170	39
2~4	6	10
4~8	1	10
8~16	0	0
16~32	0	0
32~64	0	20
64~128	0	8
128~	0	19

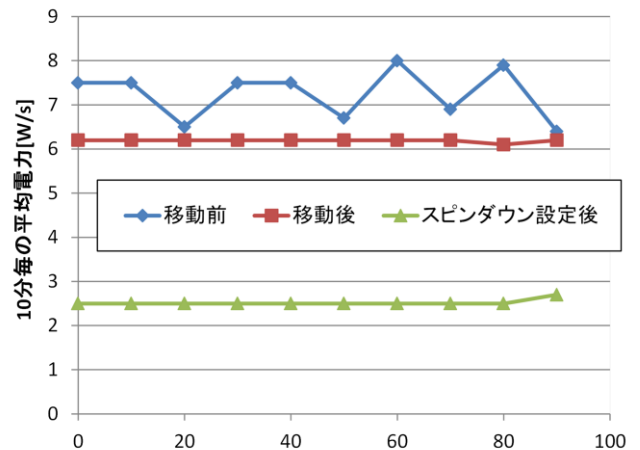


図 7 移動前後の 10 分毎の平均電力推移

んど見られなく、低下幅は平均でも 1% 未満と移動後と変わらない値となっている。

ロングインターバルのあるテーブルデータを配置した

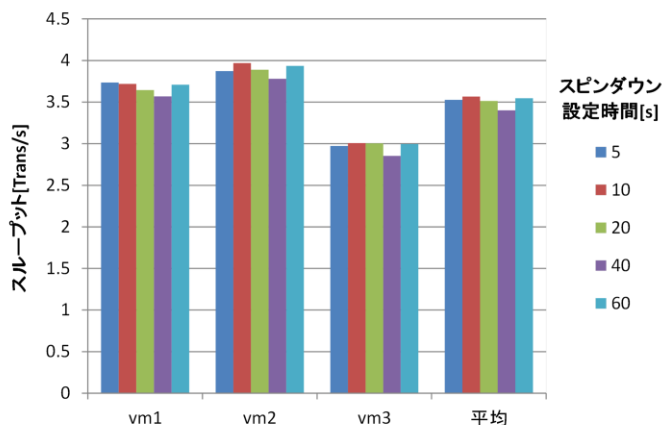


図 8 スピンドアウン設定時間によるスループット

HDD における移動前後とスピンドアウン時間を設定したときの消費電力の結果を図 6 と図 7 に示す。図 6 より、移動後は移動前よりアクセス数が減少したため、移動前より消費電力が約 14%低下している。移動かつスピンドアウン時間設定後は多くの停止時間が確保されているため、移動前と比べて 65%ほど消費電力を下げることができている。また移動後と比べても 60%ほど消費電力を下げることもできるので、暇な HDD を停止するだけでも十分な効果があることが分かる。HDD は 3 台使用しているため、移動前と比べて全体では約 22%消費電力を削減できている。図 7 より、アクセス間隔の長い移動後やスピンドアウン設定後は 10 分毎の推移も緩やかだが、移動前のアクセス間隔の短い時は推移が激しく揺れていることがわかる。

スピンドアウン設定時間によるスループットと平均消費電力の関係を図 8 と図 9 と図 10 に示す。図 8 よりスピンドアウンの設定時間によるスループットの大きな変化は見られなかった。しかし図 9 より平均消費電力では、スピンドアウン設定時間が 5 秒と 10 秒では平均消費電力が 40%ほどと 60%ほど増加してしまった。図 10 より、スピンドアウン設定時間が 5 秒のときは停止後すぐにアクセスが発生したために消費電力が増えており、9 秒のときは停止時間前にアクセスが来てしまったために停止できなくて消費電力を減らすことができなくなっている。またスピンドアウン設定時間が 5 秒のときはスピンドアウンしてからスピンドアアップしているため、設定時間が 9 秒よりも大きな消費電力が発生している。

## 5. まとめ

本稿では、アプリケーションの動作情報を用いてディスク上のデータレイアウトを変更し、HDD の消費電力を削減する手法を紹介した。そして、その手法を仮想化環境に適用し、評価結果を示した。結果から、アクセス数の多いデータを集中配置したためスループットが減少したが、アク

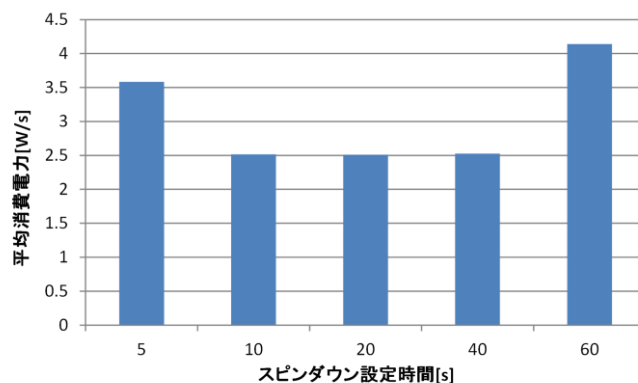


図 9 スピンドアウン設定時間による消費電力

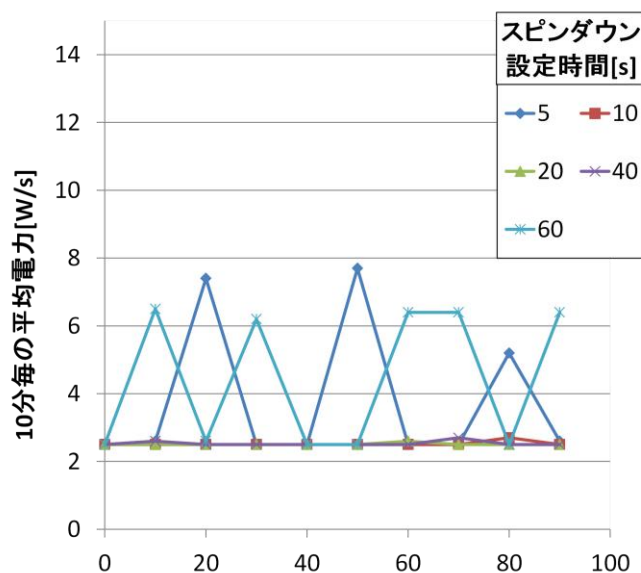


図 10 スピンドアウン設定時の 10 分毎の平均電力推移

セス間隔の拡大により大幅な省電力化ができることがわかった。またスピンドアウンの設定時間によってスループットに大きな差は見られなかったが、設定時間を適切に設定しないと消費電力が増えてしまうことがわかった。

今後は、メモリの使用について調査し、その効果の検証を行っていく予定である。

## 謝辞

本研究は JSPS 科研費 24300034, 25280022, 26730040 の助成を受けたものである。

## 参考文献

[1] Norifumi Nishikawa, Miyuki Nakano and Masaru Kitsuregawa, "Energy Efficient Storage Management

Cooperated with Large Data Intensive Applications,” 28th IEEE International Conference on Data Engineering (IEEE ICDE 2012),

[2] Norifumi Nisikawa , Miyuki Nakano and Masaru Kitsuregawa, ” Energy Efficient Storage Management Cooperated wuth Large Data Intensive Applications,”

28th IEEE International Conference on Data Engineering (IEEE ICDE 2012),

[3]西川 記史, 中野 美由紀, 喜連川 優” アプリケーション処理の I/O 挙動特性を利用したディスクの実行時省電力手法とその評価:オンライントランザクション処理における省電力効果” 電子情報通信学会論文誌, J95-D, 3, 1-13 (2012.03)